

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:01 ; Search time 8358.86 Seconds  
(without alignments)  
12307.996 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_11872\_16104

Perfect score: 4233  
Sequence: 1 atgacgacatcagctctc.....agggcaggagacgtacgctaa 4233

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estc3a.\*  
2: em\_estc3b.\*  
3: em\_estc3c.\*  
4: em\_estc3d.\*  
5: em\_estc3e.\*  
6: em\_estc3f.\*  
7: em\_estc3g.\*  
8: em\_estc3h.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_est3.\*  
12: gb\_est4.\*  
13: gb\_est5.\*  
14: gb\_est6.\*  
15: em\_estfun.\*  
16: em\_estfun.\*  
17: em\_ges\_hum.\*  
18: em\_ges\_inv.\*  
19: em\_ges\_pln.\*  
20: em\_ges\_vrt.\*  
21: em\_ges\_fun.\*  
22: em\_ges\_mam.\*  
23: em\_ges\_mus.\*  
24: em\_ges\_pro.\*  
25: em\_ges\_rtd.\*  
26: em\_ges\_phg.\*  
27: em\_ges\_vrl.\*  
28: gb\_gsa1.\*  
29: gb\_gsa2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175.4	4.1	812	29	B2671349 PUBKS59TD
2	174.4	4.1	627	29	B2563956 pac62-164
3	157.6	3.7	1552	29	B2572480 msh2_2654
4	133	3.1	765	29	B2562842 pac62-164

C	5	128.8	3.0	646	28	A0989512
6	125	3.0	1117	29	B2548527	
7	121.6	2.9	921	29	B2570019	
8	120.8	2.9	1038	29	B2567332	
9	119	2.8	667	29	B2548604	
10	117	2.8	753	29	B2561650	
11	114.4	2.7	840	29	B2558891	
12	114	2.7	827	29	B2680689	
13	111.8	2.6	901	29	B2574671	
14	110.4	2.6	889	29	B2551143	
15	109.2	2.6	633	29	B2551440	
16	107.8	2.5	827	29	B2559631	
17	107.6	2.5	648	29	B2548537	
18	107	2.5	709	28	BH488170	
19	104	2.5	1429	29	B2577585	
20	101.6	2.4	688	29	B2567972	
21	98.6	2.3	792	29	B2574700	
22	97	2.3	780	29	B2573398	
23	96.2	2.3	1342	29	B2556627	
24	96	2.3	1122	29	B2558365	
25	94.2	2.2	778	29	B2573262	
26	93.4	2.2	915	29	B2568304	
27	93.2	2.2	1307	29	B2576975	
28	92.8	2.2	628	12	B1180727	
29	92.2	2.2	580	29	B2561162	
30	91.6	2.2	885	29	B2551313	
31	91.6	2.2	1409	29	B2577581	
32	90.6	2.1	610	29	P974L	
33	90.2	2.1	1388	29	B2555697	
34	89.6	2.1	1118	29	B2551441	
35	89.6	2.1	1150	29	B2578915	
36	89	2.1	891	29	B2567106	
37	88.8	2.1	850	29	B2670377	
38	87.8	2.1	1131	29	B2559111	
39	87	2.1	1489	29	B2572227	
40	85.6	2.0	1154	29	B2553289	
41	84.8	2.0	1364	29	B2567613	
42	84.2	2.0	1594	29	B2551753	
43	83.8	2.0	715	29	B2560475	
44	83.8	2.0	807	29	B2568590	
45	82.8	2.0	881	29	B2549419	

#### ALIGNMENTS

RESULT 1  
LOCUS B2671349 812 bp DNA linear GSS 05-FEB-2003  
DEFINITION PUBKS59TD ZM 0.6 1.0 KB Zea mays genomic clone ZMBR070721,  
genomic survey sequence.

ACCESSION B2671349  
VERSION B2671349.1 GI:28219323  
KEYWORDS GSS.

SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

AUTHORS Whitehead, C.A., Quackenbush, J., Van Aken, S., Uterback, T., Resnick  
, A., Frazer, C.M., Yuan, Y., San Miguel, P., Ma, J., and Bennett, J.

TITLE Maize Genomics Consortium  
JOURNAL Unpublished  
COMMENT Contact: Cathy Whitehead  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208

Email: whitehead@tigr.org  
Seq primer: TP  
Class: sheared ends.

FEATURES Location/Qualifiers

```

source
1..812
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBA077321"
/clone_1ib="ZM_0.6_1.0 KB"
/notes="Vector: PCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"
BASE COUNT      155 a      255 c      258 g      144 t
ORIGIN

Query Match      4.1%; Score 175.4; DB 29; Length 812;
Best Local Similarity 52.7%; Pred. No. 2.1e-26;
Matches 433; Conservative 0; Mismatches 376; Indels 12; Gaps 2;

1651 TACGAGAGCTTTCGCGCGTTGCGCGACCTTGGCGCGCGCGCGCGAGAGAGGAGA 1710
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
812 TAGGCCCAATTGAATAGCCGCGCAACCGCTAGGCCCATCGCTGATGGAAGCGCGCTT 753
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1711 CGCCGGAACACATTGTCGCGGTGTGTATGAGAGAAAGCTGGAGAGAGTTTCTCGGTT 1770
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
752 GGGCCGGAAGTACTGGTGGGGCTGGCGGTGAGCGCTCCATCGACATGTTGGTGGCGCTG 693
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1771 CTGCGGTCTGAGTCAAGCGCGCTTACGTGCGCATGATGCCGACCTACCGGCGAG 1830
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
692 CTGCGGTCTCAAGCGCGGTGCGCTTATGTGCGCTGACCGGAAATACCGCGTGAAG 633
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1831 CGATCCATCTACCTCTCGATCATGTGAGTAAAGCTGTGCTGACGAGCCATGGCTG 1890
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
632 CGCTTGCTTACATGCTGGAAGACATGGCGTGAACCTGCTGACCCAGGCGCATCTG 573
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1891 GATGCAACCTGTATGCGCGCGCGGATTCAGCGGCTGCTGTAAGCAAGCGCGCTG 1950
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
572 CGTGAACATGTTGCGCATTCCTCCAGGCTGGAACCTGTGTTGGTATGATCCGATTAC 513
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1951 GAAGGAGAGGAGACAGCGCTCGATGATGCCATTCAGACCTTCGGATCTCGGCTAT 2010
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
512 GCGGCTTACAGGAGCCCAACCCAGGTATTCCTCC--CTGAGTGGCAAAACCTGGCTAT 456
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2011 GTATATACACCTCGGATTCACAGGTTGCCCCAGGGGTGATGATCATCTCGGGT 2070
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 GTGATTAATACCTCGGCTTCACCGGCGACGCTTAAAGGCGGCGGACCGTATTTGGGG 396
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2071 GCGGTCAACACATCTCTGATCAACAGAGCGCTTGAATAGGGCCCGAGACAGGGTG 2130
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
395 TTGCTCAACCGCTTGCATGATGACAGAAAGCTATGCGCTGGGCGCCAGCGCATCGGTG 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2131 CTGGGCTCTCTCGCTGAGCTTCGATCTCGGCTGATGATGTTGTTGGGATCCTGGGG 2190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
335 CTGCAAAAGACCCGTTCAAGCTTCAGCTGCGGTGGGAGTTTCTGCGCGCTGATG 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2191 GGGGCGGTACGATCTGTGTGCGGACGCGCTCCAGCTGCGCGATCGCGGCGATTGGGGA 2250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 ACCGGGCGCGCTGATGCTGCTGCGCGCGGTGATCACCGTATCCGCGCAAGCTGATC 216
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2251 GAGTTGATGAACGAGAGAGTGAAGCTGATCGGTGCGCGCTGATGCGGATG 2310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 AACCTGATCAATCAACAGAGTCAACAGCTGACCTTGTGCGGTGATGTTGCGAGGCG 156
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2311 CTGCTGAGCATTTTGAAGGTGCGCGGATTCGCTGCTAGCTCTGCGGCGCTTTGGCTG 2370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 TTC-----CTGAGGATTCGGCAGTGTGACCTGCGCAGAGCTTGCACCAACCATGTTG 105
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2371 CTGAGCGGCGATGATCCCGGTGGGCTGCTGCGGAGCTCCAGGCAATCAGGCCCGCG 2430
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 TGAAGGGCGCAACCTTGCCTGTGATGCCAGACAGAGTCTTTCGCAAGCTGCCGAC 45
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2431 GTGTGATGATGAGCTTGGCGGGGCGACCGAAGCTGAT 2471
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 GCGGCGCTTCAACCTGTACGGCGCGACCGAAGCGGCGCAT 4
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 2
B2563956
LOCUS      B2563956      627 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION pac62-164_4474.x1 pac62-164 Pseudomonas aeruginosa genomic clone
ACCESSION pac62-164_4474, genomic survey sequence.
B2563956
VERSION    B2563956.1 GI:27188838
KEYWORDS   GSS.
SOURCE     Pseudomonas aeruginosa
ORGANISM   Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 627)
            Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
            Burns, J.L., Kaul, R. and Olsen, M.V.
            Whole-Genome-Sequence Variation among Multiple Isolates of
            Pseudomonas aeruginosa Library
            J. Bacteriol., (2002) In press
            Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: ckraymond@u.washington.edu
            Class: shotgun.
FEATURES
source
1..627
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac62-164_4474"
/clone_1ib="pac62-164"
/notes="Clinical isolate 2-164 Whole genomic shotgun
library."
BASE COUNT      103 a      220 c      199 g      104 t      1 others
ORIGIN

Query Match      4.1%; Score 174.4; DB 29; Length 627;
Best Local Similarity 60.9%; Pred. No. 3.2e-26;
Matches 300; Conservative 0; Mismatches 192; Indels 1; Gaps 1;

1997 CGGATCTGCGGTATCTATCTACCTCGGATTCACAGGGGTTCGCAAGGGGGTATGA 2056
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 CGAGCGCGCTCCGTGAGCTACAGTCCGCTCCACCGCAGCCAAAGGCGTATGTC 176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2057 TCGATCATGCGGGTCCCGTCAACACCATCTGGAATCAAGAGCGCTTGAATAGGCG 2116
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
177 TACGCTACCGGGGTGAGCAACAGCTGCTGACATCAACAGCGCTACCGCGTCAAG 236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2117 CCGAGACAGGCTGCGGCTCTCTCGCTGAGCTTGGATTTCTCGGTCTATGATGCT 2176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 CCAAGACCGGCTCTCGGCTCGCGCTGAGCTGAGCTTGCATCTCGGTCTACGACTTCT 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2177 TGGGATCTGCGCGCGCGCGGTGAGATCTGTGTCGCGAGCGGCTCAAGCTGCGGATC 2236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 TCGGGGCGACCGCGCGGTGGGCCAGGTGACCTTCGACACCTGCGCGGAACGATC 356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2237 CGGCGCATTTGGCAGATTGATGACAGAGAGAGTGAAGCTGTGAACTTCGTTGCGCG 2296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
357 CATCTGACGGGCTGTGCTGCTGGAACGCAAGCATACCTGGGAAACTGGTGGCGG 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2297 CGCTAGCGGATGCTGTGAGCATTTTGAAGGTGCGCGCGATTCGCTGCTAGTCTC 2356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 CCAAGGCGCAGATGCTATCGATTACCTGGAAGGAGAGCGCGCAAGCTGCGCGGAC 476
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2357 TCGGCTTTCGCTGCTGAGCGCGCGATCGGATCCGCGGGCGCTGCGCGAGCTCCAG 2416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 CGAGCTGCTGTGCTGCTCGGAGTGAATTCGCTGACGCTGCTC-GAGCGCTGTGGC 535
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2417 CCATGAGCGCGCGCTGTGATGATCAGCTTGGGCGGCGCGACCGAAGCTGATCTGCT 2476
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



Db 536 GGTGTGCGCCGAGACGCTGCTGTAGCTGGCGAGCAGCCAGCGCGGATCTTGGC 595  
OY 2477 CCATCGGGTACCC 2489  
Db 596 CGATCGAAGATCC 608

RESULT 3  
B2572480/c  
LOCUS B2572480 1552 bp DNA linear GSS 17-DEC-2002  
DEFINITION msh2\_2654.x1 msh Pseudomonas aeruginosa genomic clone msh2\_2654,  
genomic survey sequence.  
ACCESSION B2572480  
VERSION B2572480.1 GI:27207541  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
REFERENCE Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 1552)  
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence Variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES  
source location/Qualifiers  
1..1552  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="MSH"  
/db\_xref="taxon:287"  
/clone="msh2\_2654"  
/clone\_lib="msh"  
/note="Environmental isolate. Whole genomic shotgun  
library."

BASE COUNT 320 a 448 c 387 g 390 t 7 others  
ORIGIN

Query Match 3.7%; Score 157.6; DB 29; Length 1552;  
Best Local Similarity 54.4%; Pred. No. 1.3e-22;  
Matches 356; Conservative 0; Mismatches 275; Indels 24; Gaps 1;

OY 1653 CGAAGAGCTTTCGCGCGCTTGGCGCGACCTGGCGCGCGCGCGCGAGAGCGGCGACG 1712  
Db 724 CGAGCAACTGTCCGAGCTGTGCACTGCGATCCCGCGCGCTTGCACGAGCCGCGGTGCG 665  
OY 1713 CCCGAAACATTTGATGCGCGGTGATGAGAAAGCTGGAGAGAGTTTGTGCGCGTTCT 1772  
Db 664 CCTTGGCGAAGCGGTGAGGTCAAGCTCCCGCGCGAGCCGAGAGGTTCGCGCGGTATT 605  
OY 1773 CGCGGTCTGAGTCAGCGCGCGCTTACGTGCGGATGATGCGGACTTACCGCGGAGCG 1832  
Db 604 CGGCGTCTCGCGCGAGCGCGCTGCTACGTGCGGCTGAGACACGAGCGCGCGCGACG 545  
OY 1833 TATCCACTACCTCTCGCATGATGAGGTAAAGTCTGCTGAGAGCGAGCCATGCTGGA 1892  
Db 544 CGCGGCTGTGATCGAAGAGCGCGCGGATGCTGCGGATCACCGAGAGAGAGATCC 485  
OY 1893 TGGCAAACTGTATGCGCGCGCGGAGATCAGCGCGTGTCTGTGAGCGAGCGCGGTGGA 1952  
Db 484 GCAAGCCTTGTGCGCGCGCGGTGAGTTCAGGCGCTGTGCGCGCGCGCGCGCGCGCGCG 425  
OY 1953 AGGCGAGCGGACGAGCTTCGATGATGCCATTCAGACACTTGTGAGATCTTCGCTATGT 2012  
Db 424 CCCCGTGTCCG-----CTGGCGCGCGAGGCGAGGTGCTATGT 389

OY 2013 CATCTACACCTCGGATCCACAGGGTTGCCCAAGGGGTGATGATCATTCGGGGTGC 2072  
Db 388 GATCTACACCTCGGGCTTCACCGGGGTGCCAAGGGGTGAGGTCAAGCCAGCGGGCGGC 329  
OY 2073 GTTCAACATCATCTGACATCAACAGCGCTTCAATATGAGGCCCGGAGACAGGTTCT 2132  
Db 328 GATCAATACATTCAGACCGCGCTGCTGACCTGCTCGGGTGAACCATGATGATCTTGCT 269  
OY 2133 GGGCGCTCTCCGTGAGCTTGTGATCTCTGCTGATGATGATGATGATGATGATGATGATG 2192  
Db 268 GGGCGTCTCGCGCTGACCTTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 209  
OY 2193 GGGCGTACGATCTGTGTGCGGAGCGGTCCAGCTGCGGATCCGCGGCTTGGGCGAGA 2252  
Db 208 CGGTGCGAGCTGTGCTGCTGCGGCGCCAGAGAACGCGCGGATCCGCTGTGCGGCGGA 149  
OY 2253 GTTATGACACGAGAGAGGTGACGCTGTGAGTCTGCTGCGCGCGCTGATGCGG 2307  
Db 148 GGTATTCAGCGCGCATGCGGTGAGCTGTGAGTCTGCGCGCGCGCGCTTGTGCGG 94

RESULT 4  
B2562842/c  
LOCUS B2562842 765 bp DNA linear GSS 17-DEC-2002  
DEFINITION pacs2-164\_3957.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
pacs2-164\_3957, genomic survey sequence.  
ACCESSION B2562842  
VERSION B2562842.1 GI:27185526  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
1 (bases 1 to 765)  
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence Variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES  
source location/Qualifiers  
1..765  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacs2-164\_3957"  
/clone\_lib="pacs2-164"  
/note="Clinical isolate 2-164 Whole genomic shotgun  
library."

BASE COUNT 157 a 234 c 248 g 125 t 1 others  
ORIGIN

Query Match 3.1%; Score 133; DB 29; Length 765;  
Best Local Similarity 52.0%; Pred. No. 1.5e-17;  
Matches 323; Conservative 0; Mismatches 295; Indels 3; Gaps 1;

OY 1691 GGTTCGCGAGCAGCGGCGACGCCGACACATTTGTCGCGGTGATGAGAAAGCT 1750  
Db 673 GTTCGCGGAGAGAGGTTCGCGATCATGATGATGATGATGATGATGATGATGATGATG 614  
OY 1751 GGGAGCAGTTGTGCGGTTTTCGCGGTGCTGAGTCAAGGCGCGGCTTACTGCGGATCG 1810  
Db 613 TACCATGTGTGTGCGCTGCTGACGCTGCTCAAGACCGGTGCGGCTTATGTGCGGTG 554

FEATURES	Location/Qualifiers
source	1..646
Db	1811 ATGCCGACCTTACCGGCGGAGGGTATTCACCTACCTCTCGATCATGTGAGGAAAGCTCG 1870
Db	553 ATCCGCAATATCCCGCGGACCGCTTGCAGTACATATGACGACACGCGGACTGCGCCTCT 494
Qy	1871 TCTTGACGACAGCATAGCTGTGATGAGCAAACTGTCAATGACCGCCGCGGGATTCAGCGGCTGC 1930
Db	493 TGCTCAGCCAGCAAGCGCTGTGGCGCGCTTGCCGAGTGGATGGGCTGCAGAAAGCTGC 434
Qy	1931 TCTTGACGAGAGCGCGCGCTCCAGAGCGCACCGCGACCTTCCATGATGCCCATTC-- 1987
Db	433 TGCTTCGACGACCTTGAGAGGACCTGTGACACCGGCTACCCGCGGAAAAACCGGACCTGCGCG 374
Qy	1988 AGACACCTTCGGATTCGCGGTATGTCATCTACACTCGGAGATTCACAGAGTTTGCCCAAG 2047
Db	373 AGCGCCCGGACAGTCTTCTGTCTACGCGATCTACACTTCGGTTTCCACCGGCGACCGCAAG 314
Qy	2048 GGGTGATGATGCATGATCATCGGGGTGCGCTCAACACCATTCCTGAGCATCAACGAGCGCTTCG 2107
Db	313 GCGTATGATGGTCCGCCATGAGCTTGACCAACTTGTGTGACAGCATTCGCCCGGCAACGG 254
Qy	2108 AAATGAGGCGCCGGAACAGAGGTGCTGGCGCTTCTCTCGTGAAGCTTCGATCTCTCGTCT 2167
Db	253 GCATGCTGTGCGGGGACCGCTGCTGTGTGTGACCAACCTTCTCTTCATATCTTCGCGCC 194
Qy	2168 ATGATGTTGTGGGATTCCTGGCGGGCGGGCGGTACAGATGTGGTGTGCGGACGGGTCAACG 2227
Db	193 TGGACCTATACGTAACCGCTGGGCGCTGGGCGCCAGCATCTGTCTGCCACGCGCGACAG 134
Qy	2228 TGCGGATTCGGCGCATTTGGGCGAGAGTTGATGCAACGAGACAGGTGACGATGTTGAAC 2287
Db	133 CCCAGATATCCGAGGCGCCTGCTGCATCTGTGTGACGAAACGGGGGATCCACTAGTTCTAG 74
Qy	2288 CGGTGCCGCGCTGATGCGGA 2308
Db	73 AGCGCGCGCCACCGCGGTGA 53
RESULT 5	
AQ989512/c	646 bp DNA linear GSS 14-AUG-2000
LOCUS	RefC0062 Photorhabdus luminescens strain W14 M13 library
DEFINITION	Photorhabdus luminescens genomic clone PLG00062, genomic survey
ACCESSION	AQ989512
VERSION	AQ989512.1 GI:9648106
KEYWORDS	GSS.
SOURCE	Photorhabdus luminescens
ORGANISM	Photorhabdus luminescens Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.
REFERENCE	1 (bases 1 to 646) Dabern,P.J., Bowen,D., and Blatner,F.R. Photobacterium-Constant,R.H., Waterfield,N., Burtland,V., Perna,N.T., Fathorn,P.J., Bowen,D., and Blatner,F.R. A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	
COMMENT	Contact: ffreuch-Constant RH Department of Biology and Biochemistry University of Bath South Building, Bath BA2 7AY, UK Tel: (44) 1225 826621 Fax: (44) 1225 826779 Email: bsarfcbath.ac.uk This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see ffreuch-Constant et al. 2000, Nucleic Acids Res. Seq primer: M13 Forward Class: shotgun.

		/organism="Photorhabdus luminescens"	
		/mol_type="genomic DNA"	
		/strain="W14"	
		/db_xref="taxon:29488"	
		/clone="PLG00062"	
		/dev_stage="primary phase variant"	
		/clone_1fb="Photorhabdus luminescens strain W14 M13 library"	
		/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janss."	
BASE COUNT	177 a 163 c 151 g 150 t	5 others	
ORIGIN			
Query Match	3.0%;	Score 128.8;	DB 28; Length 646;
Best Local Similarity	54.7%;	Pred. No. 1.1e-16;	
Matches 256;	Conservative 0;	Mismatches 212;	Indels 0; Gaps 0;
Oy	1973 CGATGATGCCCATTCACAGACACTTCCGATCTCGGTATGTCATCTACCACTCGGGATCCA	2032	
Db	469 CAATATATGGCTAAGCCGTCGGTGACTGATCCGGTTATTATCTATACCTCTGGCTCAA	410	
Oy	2033 CAGGGTTCCCAGAGGGGTATGATGCATATCGGGGTGCCGTACACAACATCTGTGACA	2092	
Db	409 CCGGTATGCCAAAGGGGTGGTTATCAGCCACCAAGCGCATCAATACCCTGTCTGGATA	350	
Oy	2093 TCACAGAGCCCTTGAATATAGGGCCCCGAGACAGGGTGCCTGGCCCTCTCGTGAGCT	2152	
Db	349 TTAATGCCCGCATCAGGTTCAAGAATGACACGATATATGGCACTCTCCGCCCTACATT	290	
Oy	2153 TCGATCTCTCGGTATGATGATGTGTCGGGATTCCTGCGCGCGCGGTACGATCGTGTGC	2212	
Db	289 TCGACTCTCCGTTATGACATTTTCGGTATTTCTCACAGCGGAGGGCCGTGTATTAA	230	
Oy	2213 CGAACGGCTCCAGCTGCGGATCCGGATCCGGCGATGGGCGAGATTGATGAAAGAGAGG	2272	
Db	229 TTATATGAAGCAGACGTCGTGATCATCAACCTGGGAAATGTTGATCTCTGTTACAAG	170	
Oy	2273 TGACGGTGTGAATCTCGGTGCGGCGCTGATGCGGATGCTCTCGACATTTTGAGGCTC	2332	
Db	169 TMAAGATATGAAACAGGATCCGGCATGTTTCACAGATTAACCTATTTGTAAGGGA	110	
Oy	2233 GCCCGAATGCGCTGCTGAGGCTCTGCGGGCTTTTCGCTGCTGAGCGGCGCATGATCCCGG	2392	
Db	109 TCGAATGATGATGACCCGAAAGTACTTTGTAACCTGATGTTGTCCGGGACTCGATTGGTT	50	
Oy	2393 TGGGCTGCTCTGCGAGCTCCAGGCCATCAGGCCCGGCGTGTGCGTGA	2440	
Db	49 TATCACTGCTGAAACGTTACCGGGCATTCATCTGATGTTGTTGA	2	
RESULT 6			
BZ548527		1117 bp	DNA linear GSS 17-DEC-2002
LOCUS	pacsl-60_1192.gb pacsl-60 Pseudomonas aeruginosa genomic clone		
DEFINITION	pacsl-60_1192, genomic survey sequence.		
ACCESSION	BZ548527		
VERSION	BZ548527.1		GI:27152108
KEYWORDS	GSS.		
SOURCE	Pseudomonas aeruginosa		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
AUTHORS	Pseudomonadaceae; Pseudomonas.		
JOURNAL	1 (bases 1 to 1117)		
COMMENT	Spencer,D.H., Raymond,C.K., Smith,E.B., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.		
	Whole-Genome-Sequence variation among multiple isolates of		
	Pseudomonas aeruginosa library		
	J. Bacteriol., (2002) In press		
	Contact: Chris K. Raymond		
	Genome Center		
	University of Washington		
	Box 352145, Seattle, WA 98105-2145, USA		
	Tel.: 2062216954		

Fax: 2066857244  
Email: cgraymond@u.washington.edu  
Class: shotgun.  
source

Location/Qualifiers  
1. 1117  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pac61-60.1192"  
/clone.lib="pac61-60"  
/note="Clinical isolate 1-60 whole genomic shotgun library."

BASE COUNT 169 a 403 c 330 g 212 t 3 others

ORIGIN

Query Match 3.0%; Score 125; DB 29; Length 1117;  
Best Local Similarity 52.5%; Pred. No. 7.9e-16;  
Matches 323; Conservative 0; Mismatches 285; Indels 7; Gaps 2;

```

QY 845 AGATCCGCTCCGCGACACGAGCAATGCGTCCGACTCTGGGCTGATTGAAGC 904
DB 53 AGCCGCGACACCGCGGCTCTGGATTCAGCTTCCGCGCGAGAGCGGCTACTCGAGC 112
QY 905 GCGGTGTGCGGAGCGCGGCTGACCCCGACGCGGCTCACTCTGCTGATTTCCGAG 964
DB 113 GTCTTCCGCGCGACAGTGGCGGTGACCTTGTCCAGCGTGTTCGGCGCTTCGGCGTGG 172
QY 965 TGATGGGCGCTGAGAGCGCGAGCCCC -CGGTTTACGCTCAATACGCTCTTCAACGG 1023
DB 173 TCTCGCGCGCTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 232
QY 1024 CTCCCGCTTCATCCGCGCGGTGAACGATATCAACGCGGGAATTACGCTGATGCTCCG 1083
DB 223 CATGCCGAGACACCGCGCTATCGCGAGGTGATCGCGGACTTACCAACCTGTGCTGCTG 292
QY 1084 GACATCGACACCACTCGCGCAAGAGCTTCAAGACGCGGCTTAAGCGTATTCAGAGCAG 1143
DB 293 GAGTCCGAGATCGAGCGCGGCGGTGCTTTCGCGGAGCGGCTTAAGAGGCTTCAGGCGA 352
QY 1144 CTGTGGGAAAGCATGATATCTGCGACGTAAGCGGCTATGAGTTCAGCGAGAGCGCGC 1203
DB 353 CTCCAGGAGCATGACACACCGCGCATTTCCCGCGCTGAGAGTCTCTCGCGAGAGC --- 408
QY 1204 CGGATCTGGGATCAACGAGCGCGATTTCCCGGTGCTGCTACGAGCGCGCTTAAC 1263
DB 409 --CGCCCGGCGAGCGCGACGCTCGCGCGGTGCTGCGGACGAACTTGGGCGAG 466
QY 1264 CAGCAAGTCTGTGTGATCACTCTGTGCAAGAGCTCGGAACTCCGCTGTAACCAAGCA 1323
DB 467 GAGGCGCTTGTCTCGCGCGGCTTTCGCGGACGCTTTCGGGATCTCAGACATGCTCTG 526
QY 1324 CAGACTCTCAGCTGCTGTGATATCAAGCTCTACAGACAGATGGGAGCTCTGCTCTC 1383
DB 527 CAGACCCCGCAAGTCTGCTTCAGACCAACGCTTACCGGGTGGGAGACGATTTCTGCTG 586
QY 1384 GCGTGGGACATGCTGAGCGAGATGTTCCCGCGCGACCTTCTGAGAGCAATGCTGAAG 1443
DB 587 GCTTGGGACAGGCTGTGCTGCTTTCGCGGAGAGCTGCGGAAACATGTTCAAGGCC 646
QY 1444 TACGTCGTTTTTCTC 1458
DB 647 TACGTGGGCGTCTC 661

```

RESULT 7  
LOCUS BZ570019 921 bp DNA linear GSS 17-DEC-2002  
DEFINITION msh2\_1128\_Y2 msh Pseudomonas aeruginosa genomic clone msh2\_1128,  
ACCESSION BZ570019  
VERSION BZ570019.1 GI:27205080  
KEYWORDS GSS.

SOURCE  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

REFERENCE  
1 (bases 1 to 921)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE  
Whole-Genome-Sequence Variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244

COMMENT

Query Match 2.9%; Score 121.6; DB 29; Length 921;  
Best Local Similarity 52.3%; Pred. No. 3.9e-15;  
Matches 293; Conservative 0; Mismatches 264; Indels 3; Gaps 1;

FEATURES  
source  
Location/Qualifiers  
1. 921  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="MSH"  
/db\_xref="taxon:287"  
/clone="msh2\_1128"  
/clone.lib="msh"  
/note="Environmental isolate. Whole genomic shotgun library."

BASE COUNT 175 a 264 c 268 g 201 t 13 others

ORIGIN

```

QY 1574 TGCACGCGCTGTGCGCGCGCGGTGCGAGCACTGCTCCATGAGTCTGCTGCTGG 1633
DB 145 TGCACCGGCTGTTCAGAGAGAGTGCAGCGCACCGCCAGCGCGCGCTGCTGCTG 204
QY 1634 CGCGGAACCGCTACGTAAGCAAGCTTTCGCGCGCGGCTTCGCGGACTTGGCGCGCG 1693
DB 205 GCGAGAGCGCGCTGACATGCTGAGCTGAACCGCGGTGCACACCGCTTGGCGATGCCC 264
QY 1694 TGCGGAGACAGGCGGCGACCGCGCAACATTTGCTGCGGCTGATGAGAAAGCTGGG 1753
DB 265 TGATGAGCGCGGAGATCGCGCGGACCGCTGTGGGCTGAGCATGAGCTTCATCG 324
QY 1754 AGCAGTTGTGCGGCTTCTCGCGGTGCTGAGTCAAGCGCGGCTTACGTTGCCATCGATG 1813
DB 325 AGATGCTGTGCGCTGATGCGATCTCAAGGCGCGGCGGCTTACGTTGCCGATGAGC 384
QY 1814 CCGACTTACCGCGGAGGCTATCCACTACTCTCTCGATCATGATGATGATGATGATGATG 1873
DB 385 CGGAGTATCCGAGAGGCGGAGGCGGCTTACGATGAGAGAGCGGCGGCTTCACTGCTGC 444
QY 1874 TGACGACAGCAATGCTGATGAGCAAACTGTATGCGCGCGGAGATTCAGCGGCTGCTG 1933
DB 445 TCAAGCAATGCACTTAAGCTGCGCGCTGCGGAGAGTGTGCAAGGATTCAGCTGAGC 504
QY 1934 TGAAGGAGCGCGGCTCGAAGGCGAGCGGCGACAGCGCTTCGATGATGCCATTGAGAC 1993
DB 505 AGCGGATGCTGCTGAGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 561
QY 1994 CTTCGGAATCTGCGGATGATGATCTTATCACTCTGAGATTCAGAGGCTTGGCCAAAG 2053
DB 562 GCGAAGATCTTGTCTTATGATCTTACCTTCGCTTCCAGCGGCAAGGCGGAGGCGGCG 621
QY 2054 TGATGATCATCGGCGTCCGCTCAACACATCTTGAATCAACGAGCGCTTCAAGATAG 2113
DB 622 GCAACCGCATTTGCGCGCTGAGCAACCGGTTGTGCTGATGCAAGAGGCTTATGGCGT 681
QY 2114 GCGCCGAGACAGGCTGCTG 2133

```

Db 682 GCGGTTGCGACACGCTGCTG 701

RESULT 8  
BZ567332 1038 bp DNA linear GSS 17-DEC-2002  
LOCUS pac62-164\_6886.y2 pac62-164 Pseudomonas aeruginosa genomic clone  
DEFINITION pac62-164\_6886, genomic survey sequence.  
ACCESSION BZ567332 GI:27198451  
VERSION BZ567332.1 GI:27198451  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 1038)  
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES  
source  
1..1038  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pac62-164\_6886"  
/clone\_11b="pac62-164"  
/note="clinical isolate 2-164 whole genomic shotgun  
library."

BASE COUNT 187 a 330 c 304 g 216 t 1 others

ORIGIN

Query Match 2.9%; Score 120.8; DB 29; Length 1038;  
Best Local Similarity 50.0%; Pred. No. 5.9e-15;  
Matches 328; Conservative 0; Mismatches 327; Indels 1; Gaps 1;

QY 1558 CTGAGCGACATACGCTGACGCGCTGTTGCGCGCGCGGCTCAGCAGCTGCCATGCAG 1617  
Db 80 CCGGACGAGCGGCGCTGCGAGCGCTGTTGCGCGCGCGCGCGCGATGAGCGCCGAGCGG 139  
QY 1618 CTCGCCGTGCTGCGCGCGCGCGCGCGCTCAGCAGCTGCTGCGCGCGCGCGCGCG 1677  
Db 140 GTGGCATCTCTCTTCCAGGAGCGCCAGCTGACCTACGCGCGCGCTCAACGCCAGCGCAAC 199  
QY 1678 CGACCTTGGGCG 1737  
Db 200 CGCTGCGCCACCGCTGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259  
QY 1738 ATGAGAAAGGCTGGAGGAGGCTTTCGCGCGCTTCGCGCGCTGCTGAGTCAAGCGCGCGC 1797  
Db 260 GTGACG 319  
QY 1798 TACGTGCGATGATGCG 1857  
Db 320 TACGTACCGCTGACCG 379  
QY 1858 GAGGTAAAGCTGCTGCG 1917  
Db 380 GCGATGCGCTGTTGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 439  
QY 1918 ATTCAGCGCGCTGCTGCG 1977  
Db 440 ATGCGCTGCGCTGCG 499

QY 1978 ATGCCATTACAGACCTTCCGATCTCGCGTATGTCATCTACCTCGGGATCCAGAGG 2037  
Db 500 CAGTTCGGCGCGCATCCGACGAACTCCGCTACGATGATGTCACCTCGGTTCCAGCGC 559  
QY 2038 TTGCCAAGGGGGT-GATGATGATCATCGGGGTCGCTCAACACATCTCGAGATCA 2096  
Db 560 CCGTCCAGGGGGTGGCGAATAGCCGAGAGAGCGCTGTGGCGGCAATACCAGTTTCGCT 619  
QY 2097 CGAGCGCTTCGAATAGGCGCGCGGAGACAGGCTGTGGCGCTCTCTGCTGAGCTTGA 2156  
Db 620 GAGATTATTGGCATCGGCGCGGACGACCGGCGTGTGCAATTCTCACCTTCATTTGCA 679  
QY 2157 TCTCTCGGTATGATGTTGTTGCGGATCTCGCGCGCGCGCGGCTAGATGCTGTGC 2212  
Db 680 CCGGTTTCGCGAGCGCTTACCACTTCGCTTGGTGGTCTGCTGTGTGTCTGC 735

RESULT 9  
BZ548604 667 bp DNA linear GSS 17-DEC-2002  
LOCUS pac61-60\_1266.s1 pac61-60 Pseudomonas aeruginosa genomic clone  
DEFINITION pac61-60\_1266, genomic survey sequence.  
ACCESSION BZ548604  
VERSION BZ548604.1 GI:27152185  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 667)  
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES  
source  
1..667  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pac61-60\_1266"  
/clone\_11b="pac61-60"  
/note="clinical isolate 1-60 whole genomic shotgun  
library."

BASE COUNT 100 a 235 c 234 g 98 t

ORIGIN

Query Match 2.8%; Score 119; DB 29; Length 667;  
Best Local Similarity 52.1%; Pred. No. 1.3e-14;  
Matches 266; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 2438 TGATCAGCTGGGCGGCG 2497  
Db 133 TGGTCAATTTTACCGGCG 192  
QY 2498 ACGTCAGCTATCGTGGCG 2557  
Db 193 TGCAGCGCGAAGCGCTGCG 252  
QY 2558 ACGTCTCATAGAGCGCGCTCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2617  
Db 253 AGGTCTTCACCGCTTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312

OY		2618	GCGGGGTGAGCTGGCACTGGGGCTACTGTCGCAGATGAAGAAGACGGCAAGAGCTCC	26177
Db		313	GCGGATCGGCTGCGCTGCGTGGCTATTTCGGCAGACCAGAGACAGACGCCCGAGCGTTTCC	372
OY		2678	TCTGTGACCCCGAAGACCGGGAGCGCCCTCTACAAAGACCGGGCATCTGGGCGCTAACCTCG	27377
Db		373	GCGTGATTCGGAGGACGCGCGGAGCGCCGCTATTCCGACCGGGGACCGCGCTTAGCATGGCGG	432
OY		2738	CCGATGGAACATCGATGTTTCATGTGGGGCGTGAAGACACCAATCAAATCAAGTTTGGCGGATAC	27977
Db		433	AGGACGGCGTGTGTGAACAACCTCGGCTCGAGTACAGACCAAGTCAAGTTCAACGCGCTTCC	492
OY		2798	GCGTTGACCTCGGGGAAATTCAGAGAAACGCTCTCAAGTCCGATCCGAACCTAGCCACCGGG	2857
Db		493	GCATGACCTGGGTGATGATTCGCTTCGTGCTTCGGCCAGATTTTCCGGGGTGAAGCAACGCT	552
OY		2858	TGATTGTGCTCCGTGCGGAACGACGCGGGAACAAGCTCTTCTAGCCTATGTGTCGCGG	2917
Db		553	GCGCCATCTCTACACGAGAGACAGCGCGGCGCTGCGCTGTGCGCTTAAGTGGGCGGCG	612
OY		2918	AGGGCACACGAGACGCGCTGCGAGCAGCA	2948
Db		613	CGTTGCAACCGCGCTCCAGGCGCTGAAGCA	643
RESULT_10		B2561650		
LOCUS		pacs2-164_3367	753 bp	DNA linear GSS 17-DEC-2002
DEFINITION		pacs2-164_3367.x1 pacs2-164 Pseudomonas aeruginosa genomic clone		
ACCESSION		B2561650		
VERSION		B2561650.1		
KEYWORDS		GI:27182090		
SOURCE		GSS.		
ORGANISM		Pseudomonas aeruginosa		
REFERENCE		Pseudomonas aeruginosa		
AUTHORS		Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
TITLE		Pseudomonadaceae; Pseudomonas.		
JOURNAL		1 (bases 1 to 753)		
COMMENT		Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V. Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol., (2002) In press Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: ckraymond@u.washington.edu Class: shotgun.		
FEATURES		Location/Qualifiers		
source		1..753		
		/organism="Pseudomonas aeruginosa"		
		/mol_type="genomic DNA"		
		/strain="2-164"		
		/db_xref="taxon:287"		
		/clone="pacs2-164_3367"		
		/clone_1fb="pacs2-164"		
		/note="clinical isolate 2-164 whole genomic shotgun library."		
BASE COUNT		129 a 257 c 244 g 123 t		
ORIGIN				
Query Match		2.8% Score 117; DB 29; Length 753;		
Best Local Similarity		52.9%; Pred No. 3.4e-14;		
Matches 353; Conservative		0; Mismatches 295; Indels 19; Gaps 4;		
OY		2195	GCGGTACGATCGTGTGCGCGACGCGTCCAAGCTCGCGATCCGGCGCATTTGGCGCAGAGT	2254
Db		732	GCGGACGCTGTGTGTGAACCGCGCGGCGGCATTTCCGACCCGCGAAGCTGTGCCCC	673
OY		2255	TGATGAACGAGAAAGGTGACGCTGTGGAAGTCCGTCGCGCGCTGATGCGAGTCTCG	2314

Db	612	TGATCATCCCGAAGAGGGGTGATATACGCTGCACTTCGTGCGGTCAATGCTGAGGCGCTTCT	613
QY	2315	TCGAGCAATTTTGAAGAGTCCGCCGATTCGCTCGTGAAGTCTCTGCGGCTTTGCTGCTGA	2374
Db	612	TGCAGAGATGAAGAGTGCCTCTCTGACACAGC-----TTGAAAACGATCTGTTTGA	562
QY	2375	GCGCGCATGGAATCCCGTGGGCTGCTGCGCAAGCTCCAGGCCATCAGGCGCCGCGCT	2434
Db	561	GCGCGAGAGCCCTCCCGCGAAGCCCAAGCAGCAGAGTGTTCGCCAAGCTGCGCAGGCGG	502
QY	2435	CGGTGATCAGCTGAGCGAGGCGCAACCGAAGCGTGAATCTGGTCCATCGAGTACCCTGTA	2494
Db	501	GCTCTATATACCTCTATGAGCCCGACCGAGCGGCGCAT-----CGAGTCAACCACTGGA	448
QY	2495	GGAACGTGACCTATCTGTGGCGAGCAATCCCTACGCGCGTCCGCTGCGCAACGAGCT	2554
Db	447	CTGTGATGAAGAGGAGCAAGGACCGCGGTGCGGATCGGCGCGCATGCGCAACTGGGCT	388
QY	2555	TCACATGCTCTGATGAGAGCGCTCGAACCGCGCTGCTGGTTCGAGGCGACTTATCA	2614
Db	387	GCTACATCTCTGATGAGCAACTGAGCGCGATGCGGTGGGGGTCTCGGCAAGTGTAC	328
QY	2615	TTGGCGGGGTGCGGCTGGCACTGGGCTACTGGCGCTAGCGCGATGAAAGAGACGCGCAAGCT	2674
Db	327	TGGCGGATCGGGGCTGGCGCTGGGCTGACCAACAGCGTCCGCGGCTGACTGCCAGCGTT	268
QY	2675	TCCTCTGTGACACC---CGAGACCGGGAGCGCTCTACAAAGACCGCGGATCTGGGCGCT	2731
Db	267	TCGTGTCCAGCCCGTTGTGTGTGCTGGGAGCGGATGTACCGCACCGCGGATTGGCGGCT	208
QY	2732	ACCTGCCGATGAGAAACATGCAATTCAATGGGCGGTAGAGACACCAATCAAGCTTGCG	2792
Db	207	ACCGCGCGAGAGGGGTGATCGAGTACCGCGGCGGATCGACCAACAGGTGAAGCTGCGCG	148
QY	2792	GATACCGCGTTAGACTCGGGGAAATCGAGG-AAACGCTCAAGTGCATCCGACGTACCG	2856
Db	147	GCTTGCGGATCGAAGCTGGGCGAGATCGAGGCAAGCTTGCTGGAGCATCTGTGGGTGCG	88
QY	2851	GACGCGG 2857	
Db	87	GAGGGG 81	
RESULT 11	BZ558891	840 bp DNA linear GSS 17-DEC-2002	
LOCUS	BZ558891		
DEFINITION	pac62-164_109.81 pac62-164 Pseudomonas aeruginosa genomic clone		
ACCESSION	pac62-164_109, genomic survey sequence.		
VERSION	BZ558891		
KEYWORDS	BZ558891.1 GI:27174215		
SOURCE	GSS.		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Pseudomonas aeruginosa		
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
TITLE	Pseudomonadaceae; Pseudomonas.		
JOURNAL	1 (bases 1 to 840)		
COMMENT	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.B., Hastings,M.,		
	Burns,J.L., Kaul,R. and Olsen,M.V.		
	Whole-Genome-Sequence variation among multiple isolates of		
	Pseudomonas aeruginosa library		
	J. Bacteriol., (2002) In press		
	Contact: Chris K. Raymond		
	Genome Center		
	University of Washington		
	Box 352145, Seattle, WA 98105-2145, USA		
	Tel: 2062216854		
	Fax: 2066857244		
	Email: craymond@u.washington.edu		
	Class: shotgun.		
FEATURES	location/Qualifiers		
source	1..840		
	/organism="Pseudomonas aeruginosa"		

```

/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_109"
/clone_1ib="pacs2-164"
/note="clinical isolate 2-164 whole genomic shotgun
library."
BASE COUNT      113 a      296 c      272 g      159 t
ORIGIN

```

```

Query Match      2.7%; Score 114.4; DB 29; Length 840;
Best Local Similarity 53.1%; Pred. No. 1.2e-13;
Matches 292; Conservative 0; Mismatches 251; Indels 7; Gaps 2;

```

```

QY 935 CGGGGCTACCTCGCTGATTTTCCGAGTATCGGGCGCTGAGACCGGCCCCGGT 994
DB 45 CGAGCGTTCGCGCTGCGCTTCGCGCTGCTGCGCGCTGAGACG-AAACGCCGAAAT 103
QY 995 TTACGCTCAACATACGCTCTTCAACGGGCTCCCGCTCATCGGCGCGTGAACGATATCA 1054
DB 104 TTCTCTCAACGCTGCGCTGCTTCAAGCGGCGATGCGACGACCGCTATCGGAGTGA 163
QY 1055 CGGGGACTTCAAGTATGATGCTCTCTGACATCGACACCACTCGCGCAAGAGCTTCG 1114
DB 164 TGGCGACTTCAACGCTGCTGCTGCTGAGTGCAGAGCGCGGAGTGTCTCTCG 223
QY 1115 AACAGCGGCTAAGCGTATTCAGAGCAGCTGTGGAGCGATGATCACTGCGACGTAA 1174
DB 224 CGAGGCGGTGAAGAGCTTCAAGCGCACTTCAAGAGCAGACCACTGACGCGCATTC 283
QY 1175 GGGGTATGAGTCCAGAGGAGCGCGCGCGCTCTGAGGATTCAGAGAGGCGCATTC 1234
DB 284 CGGCTTGAAGGCTGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 337
QY 1235 TCCCGTGTGCTCAAGAGCGCGCTTCAACGCAAGTCTGTGCTCACTGCTTGCAGA 1294
DB 338 CGGTGTGTGCTGCGAGCAACCTGAGGAGAGGAGGCTTCCGCGGCGCTTCCGAGG 397
QY 1295 GCGTGGAACTCGGTGTAACACGAGCGGCTTCAAGCTGCTGCTGATCATCAGC 1354
DB 398 CTTTCGGGATCTCCAGCATGCTCTCCAGACCGCGAGGCTGCTGCTGACACACAG 457
QY 1355 TCTAGAGACATGAGGAGCGCTGCTGCGGAGCACTGCGAGGAGTGTCCGCG 1414
DB 458 TCTAGCGGTGAGCGAGGATCTGCTGCGGAGCGGCTGCTGCTGCTGCTGCTGCG 517
QY 1415 CCGACCTTGAACGATGCTGAGAGCGTACGCTTTTCTCCGCGGCTCATCTAGG 1474
DB 518 AAGGTGCGCGGAAACCATGTTGAGAGCTTACGTGGGCTGCTCCAGCGTCTCTGAGA 577
QY 1475 AACCATGGG 1484
DB 578 CGCTGGGAG 587

```

```

RESULT 12
BZ680689      827 bp      DNA      linear      GSS 05-FEB-2003
LOCUS      BZ680689
DEFINITION      PUBEL33FD ZM.0.6.1.0 KB Zea mays genomic clone ZMBR1036F18,
genomic survey sequence.
ACCESSION      BZ680689
VERSION      BZ680689.1 GI:28235748
KEYWORDS      GSS.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
Clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 827)
AUTHORS      Whitefaw,C.A., Quackenbush,J., Van Aken,S., Uteback,T., Resnick
A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE      Maize Genomics Consortium
JOURNAL      Unpublished

```

```

COMMENT      Contact: Cathy Whitefaw
TIGR      9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitefaw@tigr.org
Seq primer: 7F
Class: sheared ends.
FEATURES
source
1..827
/mol_type="genomic DNA"
/strain="873"
/db_xref="taxon:4577"
/clone="ZMBR1036F18"
/clone_1ib="ZM.0.6.1.0 KB"
/note="vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"
BASE COUNT      147 a      294 c      241 g      145 t
ORIGIN

```

```

Query Match      2.7%; Score 114; DB 29; Length 827;
Best Local Similarity 50.8%; Pred. No. 1.5e-13;
Matches 301; Conservative 0; Mismatches 285; Indels 6; Gaps 1;

```

```

QY 569 AGGACCGGAAACCCGCTCTGCTCAGTATGATCTATTACGTTACGACGACGCC 628
DB 241 AGGCGCGGACACAGCGCTGAGTACGCTGAGACAAATTGATCTCGACGCTTGA 300
QY 629 TGTCCATCATCTTCAAGAGCTGCTCAGCTTCTCAAGAGATCCGAGACTTCTCT 688
DB 301 TCTCGAGTTTCAAGCGGATCGATGATGCTTCTACGCGGAGCGGACGACATAG 360
QY 689 TCTGAGCTCTGTAACCGCGATTAATGATCTGCGCTGAGTCTCGCAAGAGTTGAG 748
DB 361 CGTTGGGTCTGCTTGGCGATTA-----CCAATGCAAGCGCGCGGACCTGAG 414
QY 749 CGCATCAAGATGATGATTAATGAGAGCGGCGATCGCGAGCTCCACCTCGCGGA 808
DB 415 AGCTCAGCGCGGAGGCTTCTGCAAGAGCGCTGCGCAATGCGCGCGCGCGCG 474
QY 809 CGCTTCGATGAAGCGCGATCTTACCTGAGAGATCGCTTCCGCGCACAGCGAGC 868
DB 475 AGTTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
QY 869 AATGCTGCGTGAATCTCTGCGGTGATTAAGCGCGGTCTGCGGAGCGCGGCTGA 928
DB 535 GGCAGATGATGCGCGCGCTGCGAGGATCTCTCAAGGCGCGCAAGCGGAGCTCA 594
QY 929 CCGCGAGCGCGTACCTGCGGTGATTTCCGAGGTGATCGGGCGTGAAGCGGAGC 988
DB 595 CGGCTGCGGATGCTGCGGTGCTTTCGGAACCTCGCGCGCTGAGTGAAGCGC 654
QY 989 CCGGTTTACGCTCAACATTAAGCTTCAACGCGCTCCCGCTCATCGCGCGTGAAG 1048
DB 655 CGGCGCTAGCTTAACCTACATGTTGATGCGCGCGCTTACACCTCGATGAGC 714
QY 1049 ATATACCGGAGCTTCAAGTGTGCTCTGAGCATGACACCACTCGCGCAAGA 1108
DB 715 AGGTATGAGGCGATTCATGCTGACCTGCTCGGTATGTCGCGGTGAGTGAAG 774
QY 1109 GCTTGAACAGCGCGCTAAGGCTATTCAAGGACAGCTGAGGAGCGATGA 1160
DB 775 GCTGGGTGACCGTGAAGAGCATGACCAATCTCTGGCGAGGCGTTGA 826

```

```

RESULT 13
BZ574671      901 bp      DNA      linear      GSS 17-DEC-2002
LOCUS      BZ574671
DEFINITION      msh2_3778.x1 msh Pseudomonas aeruginosa genomic clone msh2_3778,
genomic survey sequence.
ACCESSION      BZ574671
VERSION      BZ574671.1 GI:27209732

```

KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
AUTHORS 1 (bases 1 to 901)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R., and Olsen, M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
JOURNAL Contact: Chris K. Raymond  
COMMENT Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: Shotgun.

FEATURES  
source Location/Qualifiers  
1..901  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="MSH"  
/db\_xref="taxon:287"  
/clone="msb2\_3778"  
/clone\_11b="msb"  
/note="Environmental isolate. whole genomic shotgun  
library."

BASE COUNT 140 a 305 c 314 g 140 t 2 others  
ORIGIN

Query Match 2.6%; Score 111.8; DB 29; Length 901;  
Best Local Similarity 52.4%; Pred. No. 4.3e-13;  
Matches 273; Conservative 0; Mismatches 242; Indels 6; Gaps 1;

QY 1536 GCAGAAACGCGACCAAGCGCTGTGAGGAGATACGCTGACGCGCTTTGCGCGCG 1595  
DB 601 CGCGGCGACGCGCGCGCGAGGAGGCTCCAGAGACCTCCATGCGCTTTGCGCGCG 542  
QY 1596 GGTGAGAGAGCTCCCATGACAGCTGCGCTGTGTGTCGCGCGGAGAGAGCTCAGTACGA 1655  
DB 541 CGTGGCGGCGACCGCGAGCGCGCGCGCTGACCTTCCCGGACGAGACCTGAGCTATGC 482  
QY 1656 AGAGCTTTCGCGCGCTTGCAGCGCTTGTGCGCGCGCGCGCGCGAGCGCGCGCGCG 1715  
DB 481 CGAAGCTGATGCGCGCTTCAACCGCTGTGCGCGCGCTTGTGCGAGTACAGCGCTGCGCG 422  
QY 1716 GAACACATGTGTGCGGTGTGTGATGAGAAAGCTGGAGAGAGTGTGTGCGCGTTCGC 1775  
DB 421 GGAAGTGGGGTTCGCGCTTGTGCGCGCTTGTGCGCGCTTGTGCGCGCTTGTGCGCG 362  
QY 1776 GGTGCTGAGTACGCGCGCGCGCTTACGCTGCGCGCTGCGCGCGCGCGCGCGCGTAT 1835  
DB 361 GATCTCTCAAGCGCGCGCGCGCTTACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCT 302  
QY 1836 CCAGTACTCTCCATCATGTGATGAGTAAAGCTGTGCTGACGAGCGCGCGCGCGCGCG 1895  
DB 301 GCAATCATGATTCAGGAGACAGGCGCGTACGCTGTCTCAAGCGCGCGCGCGCGCGTTCGA 242  
QY 1896 CAAGTGTATGCGT 1949  
DB 241 GCGCGTGGCGAGTGGCG 182  
QY 1950 CGAAGCGAGCG 2009  
DB 181 GCTGACCGCGAGAGCG 122  
QY 2010 TGTGATTCACCTGCGGATCCAGAGGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 2050  
DB 121 CTTGATTCATCTCGCGCTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81

RESULT 14  
B2551143/c 889 bp DNA linear GSS 17-DEC-2002  
LOCUS pasci-60\_3008.x1 pasci-60 Pseudomonas aeruginosa genomic clone  
DEFINITION pasci-60\_3008, genomic survey sequence.  
ACCESSION B2551143  
VERSION B2551143.1 GI:27154724  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
AUTHORS 1 (bases 1 to 889)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R., and Olsen, M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
JOURNAL Contact: Chris K. Raymond  
COMMENT Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: Shotgun.

FEATURES  
source Location/Qualifiers  
1..889  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pasci-60\_3008"  
/clone\_11b="pasci-60"  
/note="Clinical isolate 1-60 whole genomic shotgun  
library."

BASE COUNT 164 a 294 c 280 g 148 t 3 others  
ORIGIN

Query Match 2.6%; Score 110.4; DB 29; Length 889;  
Best Local Similarity 50.8%; Pred. No. 8.5e-13;  
Matches 349; Conservative 0; Mismatches 322; Indels 16; Gaps 3;

QY 2002 CTCGCGTATGATCATCAACCTCGGATCCACAGGTTGCCCAAGGGGATGATCAT 2061  
DB 800 CTGCGTATGCTTCTTCACTCGCGCTCCACCTTAAGCCAGGGGCGCGCAACCGC 741  
QY 2062 CATCGGGGTGCGCTCAACCAACCATCTGAGACATCAACGAGCGCTTCAAAATAGGCGCGGA 2121  
DB 740 CATTGGGGGTAAACCAACGCGTGTGATGACAGAGAGCTTATGGGCTGGGCGTGGC 681  
QY 2122 GACGAGGTGCTGGGCGCTCTCTGCGTGAAGCTTGCATCTCTGCTGTATGATGTTCGGG 2181  
DB 680 ACCACGGGTGTCAGAAAGACCCCGTTCAGCTTCACAGTGTGCTGGAGTCTTTCGG 621  
QY 2182 ATCTGGCGGGCGGGCGGAGG-ATCGTGTGTCGGAGCGCGTCAAGTGGCGGATTCGGCG 2240  
DB 620 GCGCGTATGAGCGGGGCGCGCTTGTGTGTACCGCGCGCGCGCGCGCGCGCGCGCG 561  
QY 2241 GCATTGGGACAGTGTGATTCAGACAGAGAAAGTACGCGTGTGAATTCGCGCGCGCT 2300  
DB 560 GAAGCTGTGTGCGCTGTATCATATCGGAAGGGGTGACACGCTGTGACATTCGTCGCGTAT 501  
QY 2301 GATCGGATGCTGTCGAGCATTTTGAAGGTTCGCGCGCGCGCGCGCGCGCGCGCG 2360  
DB 500 GCTGACAGCGCTTCTTGACAGGACGAAGAGTGCCTCTTGACACAGC-----CTGAA 450  
QY 2361 GCTTTCGCTGTGAGCGCGAGCTGATCCGCGTGGGCTGCGCGCGCGCGCGCGCGCGCAT 2420  
DB 449 ACGGATGCTTTCAGCG 390  
QY 2421 CAGCGCGCGCGCTGTGATGATGACCTGGCGGGGCGACCGAAGCGTGTGATGTGTCAT 2480

Db 389 GCTGCCGAGCGCCGCTCTATACCTTATGCCCCGACCGAGCGGCCATCGACCTCAC 330  
QY 2481 CGGGATACCCCTGAGAGACGTGACCTATCGGGGAGACATCCCTACGCGCTCGCT 2540  
Db 329 CCACTGACCTGATGAGAGAGGCAAGGACGCGGTGCCGAT-----CGCGCGCGAT 276  
QY 2541 GGGCAACGAGATTCACGCTGCTGATGAGGGGCTCGAACCGCGCCGCTGCTGCTTC 2600  
Db 275 CCGCAACCTGGGCTGTACATCTCGATGCAACCTGGAGCGGTACCGGTGGGCTGCT 216  
QY 2601 GGGGCACTCTACATTTGGCGGGGTGCGGCTGCGCATGCGCTACTGCGCGATGAAGAA 2660  
Db 215 CGCGGACGTGTACTGTGCGCGGCTGCGGGCTTGGCCCTGAGCTACCAACGCTCCGGGCT 156  
QY 2661 GACGCGCAAGAGCTTCTGTCGACCC 2687  
Db 155 GACTGCCGAGCGCTTCTGTCGCGACGCC 129

RESULT 15  
BZ551440/c 633 bp DNA linear GSS 17-DEC-2002

LOCUS pacal-60\_3153, x1 pacal-60 Pseudomonas aeruginosa genomic clone  
DEFINITION pacal-60\_3153, genomic survey sequence.

ACCESSION BZ551440.1 GI:27155021

VERSION BZ551440.1

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

REFERENCE 1 (bases 1 to 633)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
Whole-Genome-Sequence Variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Classes: Shotgun.

JOURNAL  
TITLE  
AUTHORS  
COMMENT

FEATURES  
source  
Location/Qualifiers  
1..633  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone\_lib="pacal-60\_3153"  
/note="clinical isolate 1-60 Whole genomic shotgun  
library."

BASE COUNT 97 a 226 c 214 g 96 t

Query Match 2.6%; Score 109.2; DB 29; Length 633;  
Best Local Similarity 56.1%; Pred. No. 1.4e-12;  
Matches 229; Conservative 0; Mismatches 173; Indels 6; Gaps 1;

QY 2523 CCCCTACGGCGCTGCGCTGCGCAACGATTCACGCTGCTGATGAGCGCTCGAAC 2582  
Db 554 CCGATCGGGCGCGCCCTGCGCAACGCTGCTGCCCGCTGAGACGCGAGTTCAACCT 495  
QY 2583 GCGCCCGGTCTGGGTTCCGGGGCACTTACATTGGCGGGGTGCGGCTGCGGCTA 2642  
Db 494 GCTGCGGCTGCGCGCGCGAGCTGTGATCGGGCGCTGGCCCGCGGCTA 435  
QY 2643 CTGGCGCGATGAAGAGACGCGCAAGAGCTTCTCG-----TGACCCCGAGACCG 2696  
Db 434 CTTGGGGCGCGCGCGCTGAGCGCGAGCGCTTCTGTGCGATCGTTCTCGCGCGCTGG 375

QY 2697 GAGCGCTCTTACAAAGACCGCGATCTGGCGCGCTACTGCGCCGATGGAACATCGAGTT 2756  
Db 374 CGAGCGCTGTACCGCACCGCGGACCGCGCTGGAACCGCGCGCTGTGGAATA 315  
QY 2757 CATGGGGGTGAGGACCAATCAAGCTTTCGCGATACCGCGTTGAGCTCGGGAAAT 2816  
Db 314 CTTGCGCGCGCTGCGACGAGTCACTGTGCTTCCGCAATCGAGCGGAGAAAT 255  
QY 2817 CGAGAAACGCTCAAGTGCATTCGAACTGACGCGGATTTGTGCCCGTGGAA 2876  
Db 254 CAGGCAAGCGCTGTGGCGGAGCGGGGTGGCCAGCGGATGATGATCCGGAAG 195  
QY 2877 GAGCGCGCAACAGCTCTTTCAGCTTATGTGTTCCCGAGGCGAC 2924  
Db 194 CTTGCGCGGACCGAGTTGTGCTGCTACTTACCGCGCGCTGTGGCGC 147

Search completed: October 3, 2003, 17:48:03  
JDD time : 8366.86 secs





Db 11992 ATCTCCGAGACAAAAAGACGATCTGACGATGCTCCGTACAGAGACTCCCCGACGAATCC 12051  
Qy 181 ATCTGTCGCCGCCCAAGCCGAGCGGACGCTCCGTTTCTCTCAACAGACATCCAAAGATCC 240  
Db 12052 ATCTGTCGCCGCCCAAGCCGAGCGGACGCTCCGTTTCTCTCAACAGACATCCAAAGATCC 12111  
Qy 241 TACTGCTGGGGCGGACAGAGCGTTTACGATGCCCGAGGGGATCCACGCCCTATGCGGAA 300  
Db 12112 TACTGCTGGGGCGGACAGAGCGTTTACGATGCCCGAGGGGATCCACGCCCTATGCGGAA 12171  
Qy 301 TACGACTGACGGATCTCGACGTCGCGAGGCTGAGCGCGCTTTTCGAAAAGTCGTCCG 360  
Db 12172 TACGACTGACGGATCTCGACGTCGCGAGGCTGAGCGCGCTTTTCGAAAAGTCGTCCG 12231  
Qy 361 CGGACAGACATGCTTGGGCGCACACGCTGCCCGACATGATGAGTATCGAGCTTAA 420  
Db 12232 CGGACAGACATGCTTGGGCGCACACGCTGCCCGACATGATGAGTATGAGCTTAA 12291  
Qy 421 GTGACGCGGACATCGAGATCAATCGATCTGCGCGGGCTGACCGGAGCACAGGAGG 480  
Db 12292 GTGACGCGGACATCGAGATCAATCGATCTGCGCGGGCTGACCGGAGCACAGGAGG 12351  
Qy 481 AGGCTGCTGCTGTCGAGATGAGATGTCGACCGCATCTATGACACCGAGCGCTCCG 540  
Db 12352 AGGCTGCTGCTGTCGAGATGAGATGTCGACCGCATCTATGACACCGAGCGCTCCG 12411  
Qy 541 CTCTATACGTCGTGCGCGCTTGGCTGAGACGAGCGGCAAAACCGTCTGCTCACTATC 600  
Db 12412 CTCTATACGTCGTGCGCGCTTGGCTGAGACGAGCGGCAAAACCGTCTGCTCACTATC 12471  
Qy 601 GATCTCATTAAGTTGACCTAGGACGCTGTCATCATCTTCAAGGACGGGCTCAAGCTTC 660  
Db 12472 GATCTCATTAAGTTGACCTAGGACGCTGTCATCATCTTCAAGGACGGGCTCAAGCTTC 12531  
Qy 661 TACGAGATCCGAGACCTCTCTCCCTGCTGAGCTCTGATCCGAGATTATGATCTC 720  
Db 12532 TACGAGATCCGAGACCTCTCTCCCTGCTGAGCTCTGATCCGAGATTATGATCTC 12591  
Qy 721 GCGCTGAGATCTCGCAAGAAAGTCTGAGCGGATCAACGATCGATGATTACTGAAAGCG 780  
Db 12592 GCGCTGAGATCTCGCAAGAAAGTCTGAGCGGATCAACGATCGATGATTACTGAAAGCG 12651  
Qy 781 CGCATGCGGAGACTCCACCTCGCGGAGGCTTCGATGAAGCGGATCATCTTACCTG 840  
Db 12652 CGCATGCGGAGACTCCACCTCGCGGAGGCTTCGATGAAGCGGATCATCTTACCTG 12711  
Qy 841 AAGGAGATCCGCTTCCGAGCACAGAGCAATGCTGCGATCTCGAGCTCTGGGGTCGATTG 900  
Db 12712 AAGGAGATCCGCTTCCGAGCACAGAGCAATGCTGCGATCTCGAGCTCTGGGGTCGATTG 12771  
Qy 901 AAGCGGCTGTGCGGAGAGCGCGGCTGACCCCGATCGGCGTCACTCTGGCTGATTTTC 960  
Db 12772 AAGCGGCTGTGCGGAGAGCGCGGCTGACCCCGATCGGCGTCACTCTGGCTGATTTTC 12831  
Qy 961 GAGGTGATGGGGGCTGAGAGCGGAGCCCGGTTTACGCTCAACATAAGCTTTCAAC 1020  
Db 12832 GAGGTGATGGGGGCTGAGAGCGGAGCCCGGTTTACGCTCAACATAAGCTTTCAAC 12891  
Qy 1021 CGGCTTCCCGTCCATCCGCGCTGAAAGATATCACCGGGGACTTCACTGATGATCTC 1080  
Db 12892 CGGCTTCCCGTCCATCCGCGCTGAAAGATATCACCGGGGACTTCACTGATGATCTC 12951  
Qy 1081 CTGAGCATGACACCATGCGGACAAAGAGCTTGAACAGCGGCTTAAGGTTTCAAGAG 1140  
Db 12952 CTGAGCATGACACCATGCGGACAAAGAGCTTGAACAGCGGCTTAAGGTTTCAAGAG 13011  
Qy 1141 CAGCTGTGGGAAGCGATGATATCATGCGAGCTAAAGGCTATGAGGTTCCAGGAGAGGCG 1200  
Db 13012 CAGCTGTGGGAAGCGATGATATCATGCGAGCTAAAGGCTATGAGGTTCCAGGAGAGGCG 13071  
Qy 1201 GCCCGGCTCTGGGAGTCCACAGAGGCGATTTGCTCCGCTGCTGCTCAAGCGCGCTT 1260  
Db 13072 GCCCGGCTCTGGGAGTCCACAGAGGCGATTTGCTCCGCTGCTGCTCAAGCGCGCTT 13131

Qy 1261 AACCGCAAGTGTGTGTGTCACTTCTGTCAGAGAGCTCGGAATCCCGGTGTACACAGC 1320  
Db 13132 AACCGCAAGTGTGTGTGTCACTTCTGTCAGAGAGCTCGGAATCCCGGTGTACACAGC 13191  
Qy 1321 ACGGACATCCCTCAGCTGCTGTGATCATCAAGCTCTACAGACAGATGGGAGCTCTGC 1380  
Db 13192 ACGGACATCCCTCAGCTGCTGTGATCATCAAGCTCTACAGACAGATGGGAGCTCTGC 13251  
Qy 1381 CTGCGTGGGACATGTCGACGAGTGTCCCGCCGACCTTCTGAGACATGCTCGAA 1440  
Db 13252 CTGCGTGGGACATGTCGACGAGTGTCCCGCCGACCTTCTGAGACATGCTCGAA 13311  
Qy 1441 GGTACGTCGTTTTTCTCCGCGGCTCACTAGAGAACATGGGGGTAACAGTGGCTGT 1500  
Db 13312 GGTACGTCGTTTTTCTCCGCGGCTCACTAGAGAACATGGGGGTAACAGTGGCTGT 13371  
Qy 1501 TCGCTTCGCTGCCAGCTAGAAAGCGGGGAGCGGAAACGCGACCAACGCGCTGCTG 1560  
Db 13372 TCGCTTCGCTGCCAGCTAGAAAGCGGGGAGCGGAAACGCGACCAACGCGCTGCTG 13431  
Qy 1561 AGCGACATACGCTGACAGGCTGTTGCGCGCGGCTGAGACAGCTGCCATGCACTC 1620  
Db 13432 AGCGACATACGCTGACAGGCTGTTGCGCGCGGCTGAGACAGCTGCCATGCACTC 13491  
Qy 1621 GCCGTGTGTGCGGCGGCAAGACGCTCACTACGAAGAGCTTTCGCGCGCTGCGGCGGA 1680  
Db 13492 GCCGTGTGTGCGGCGGCAAGACGCTCACTACGAAGAGCTTTCGCGCGCTGCGGCGGA 13551  
Qy 1681 CTGGCGCGGCTCGGAGACAGGGGACAGCCCGAACATTTGTCGCGTGTGTATG 1740  
Db 13552 CTGGCGCGGCTCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13611  
Qy 1741 GAGAAAGCTGGGAGCAGGTTGTGCGGTTCTCGCGTGTCTGAGTCAAGCGCGGCTTAC 1800  
Db 13612 GAGAAAGCTGGGAGCAGGTTGTGCGGTTCTCGCGTGTCTGAGTCAAGCGCGGCTTAC 13671  
Qy 1801 GTGCCGATCGATCCGACCTTACCGGCGAGAGGTAATCCACTCTCTGATGATGATG 1860  
Db 13672 GTGCCGATCGATCCGACCTTACCGGCGAGAGGTAATCCACTCTCTGATGATGATG 13731  
Qy 1861 GTAAAGCTGCTGACAGCAGCATGCTGATGAGCAACTGTCAATGAGCGCGGAGATC 1920  
Db 13732 GTAAAGCTGCTGACAGCAGCATGCTGATGAGCAACTGTCAATGAGCGCGGAGATC 13791  
Qy 1921 CAGCGCTGCTGTGAGCGAGGCGCGGCTCGAAGCGACAGGCTCTCGATGATG 1980  
Db 13792 CAGCGCTGCTGTGAGCGAGGCGCGGCTCGAAGCGACAGGCTCTCGATGATG 13851  
Qy 1981 CCCATTCAAGACCTTTCGATCTCGGATGATCTCACTCGGAGATCCACAGGTTG 2040  
Db 13852 CCCATTCAAGACCTTTCGATCTCGGATGATCTCACTCGGAGATCCACAGGTTG 13911  
Qy 2041 CCAAGAGGGGATGATGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2100  
Db 13912 CCAAGAGGGGATGATGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13971  
Qy 2101 CGCTTGAATAGGCGCCGAGAGCAGGAGCTGTGCGCTCTCTCGCTGAGCTTCAATCTC 2160  
Db 13972 CGCTTGAATAGGCGCCGAGAGCAGGAGCTGTGCGCTCTCTCGCTGAGCTTCAATCTC 14031  
Qy 2161 TCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220  
Db 14032 TCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 14091  
Qy 2221 TCCAGCTGCGGATCCGCGGATTTGGGACAGATTGATCGAAGAGAGGTGACGCTG 2280  
Db 14092 TCCAGCTGCGGATCCGCGGATTTGGGACAGATTGATCGAAGAGAGGTGACGCTG 14151  
Qy 2281 TGAAGCTGCTGCGCGGCTGATGCGATGCTCTGTCAGACATTTTGAAGGTTGCCCGAT 2340  
Db 14152 TGAAGCTGCTGCGCGGCTGATGCGATGCTCTGTCAGACATTTTGAAGGTTGCCCGAT 14211

QY	2341	TCGCTCGTAGAGTCTCTGCGGGCTTTCGTGTCGTAGACGGACGTATGATCCGGATGAGGCGTGG	2400
Db	14212	TCGCTCGCTAGAGTCTCTGCGGGCTTTCGTGTCGTAGACGGACGTATGATCCGGATGAGGCGTGG	1427
QY	2401	CTGCGAGAGCTCCAGGCCATCAGGCCCGCGGTGTCCGTGATCAGCTTGGGCGGGGCCACC	2460
Db	14272	CCTGGCGAGGCTCCAGGCCATCAGGCCCGCGGTGTCCGTGATCAGCTTGGGCGGGGCCACC	1433
QY	2461	GAACCGTTCGATCTCTGCTCATTGGGTTACCCGTAGAGAACTGTCCACTTATGTGGGGCAGC	2520
Db	14332	GAACCGTTCGATCTCTGCTCATTGGGTTACCCGTAGAGAACTGTCCACTTATGTGGGGCAGC	1439
QY	2521	ATCCCTCAACGGCCCGCTCGCTCGCAACAGACGTTTCCACGCTGCATGTAGAGGCGTGGAA	2580
Db	14392	ATCCCTCAACGGCCCGCTCGCTCGCAACAGACGTTTCCACGCTGCATGTAGAGGCGTGGAA	1445
QY	2581	CCGCGCCCGGTCTGTGGTTCCGGGGCAACTTACATTGGCGGGGTTCGGGCTTGGCATTTGGG	2640
Db	14452	CCGCGCCCGGTCTGTGGTTCCGGGGCAACTTACATTGGCGGGGTTCGGGCTTGGCATTTGGG	1451
QY	2641	TACTGGCGCGCATGTAAAGAGAGACCGCGCAGAGCTTCTCGTGACCCCGAGACCGGGGAG	2700
Db	14512	TACTGGCGCGCATGTAAAGAGAGACCGCGCAGAGCTTCTCGTGACCCCGAGACCGGGGAG	1457
QY	2701	CGGCTTACAAAGACCGCGCGATCTCTGGGGCGCTACCTGCGCCCATGTGGAAACATTCGATTCATG	2760
Db	14572	CGGCTTACAAAGACCGCGCGATCTCTGGGGCGCTACCTGCGCCCATGTGGAAACATTCGATTCATG	1463
QY	2761	GGGCGTAGAGGACCAACCAATCAAGCTTTCGCGGATACCGCGTGTAGCTCGGGGAAATCGAG	2820
Db	14632	GGGCGTAGAGGACCAACCAATCAAGCTTTCGCGGATACCGCGTGTAGCTCGGGGAAATCGAG	1469
QY	2821	GAAGCGCTCAAGTTCGATCCGACGCTACCGCGACGGGTGATTTGGTCCCGTTCGGAAACGAC	2880
Db	14692	GAAGCGCTCAAGTTCGATCCGACGCTACCGCGACGGGTGATTTGGTCCCGTTCGGAAACGAC	1475
QY	2881	GGGCGGAAACAAAGCTCCCTTCTAGCCCTATGTGTCCCGGAGGGGCACACGGACACGCGTGGC	2940
Db	14752	GGGCGGAAACAAAGCTCCCTTCTAGCCCTATGTGTCCCGGAGGGGCACACGGACACGCGTGGC	1481
QY	2941	GAGACGAGACCGGACGCTCAAGACCGGACGGGATCGACCGAGAGACACGCGCGCGAAAGCG	3000
Db	14812	GAGACGAGACCGGACGCTCAAGACCGGACGGGATCGACCGAGAGACACGCGCGCGAAAGCG	1487
QY	3001	GACGCGTTGACGACGCGCGAAGAGGGTGCAGTTCAAGCTCGCTCGACACGGAATTCGCGAGG	3060
Db	14872	GACGCGTTGACGACGCGCGAAGAGGGTGCAGTTCAAGCTCGCTCGACACGGAATTCGCGAGG	1493
QY	3061	GACCTTGCACGGAAGACCCGCTGTCTGATCTGACCGGGCAGAGATCCGGGAGAGCGGGGCTG	3120
Db	14932	GACCTTGCACGGAAGACCCGCTGTCTGATCTGACCGGGCAGAGATCCGGGAGAGCGGGGCTG	1499
QY	3121	GAGCTTACACGGCGCTCGCGCTGATGGGTCCGAAAGTTCCTTAGAGGCGCGGATTTCCGTTGTT	3180
Db	14992	GAGCTTACACGGCGCTCGCGCTGATGGGTCCGAAAGTTCCTTAGAGGCGCGGATTTCCGTTGTT	1505
QY	3181	GAGTTTGGTTCGATTCTTAGCTGCTTGAAGCAGCTGAGGCCGCGACCGGACGCACTTCTCC	3240
Db	15052	GAGTTTGGTTCGATTCTTAGCTGCTTGAAGCAGCTGAGGCCGCGACCGGACGCACTTCTCC	1511
QY	3241	AAATTCGCTTATCCATCGGCGGGCAGACGTAACCCGGTGCAAACTTACGCGCTATGTTCAA	3300
Db	15112	AAATTCGCTTATCCATCGGCGGGCAGACGTAACCCGGTGCAAACTTACGCGCTATGTTCAA	1517
QY	3301	TCGGGCGCGAATCGAGGGCGGTGAGAGAGGGCTTCTAATTATACCAACCCGTTTCGAGACCGT	3360
Db	15172	TCGGGCGCGAATCGAGGGCGGTGAGAGAGGGCTTCTAATTATACCAACCCGTTTCGAGACCGT	1523
QY	3361	TTGCTGAAGCTCTCCGATCAACGGGATCGAGCGCGGACGCACTTTCGGCAAAACTTTCGAC	3420
Db	15232	TTGCTGAAGCTCTCCGATCAACGGGATCGAGCGCGGACGCACTTTCGGCAAAACTTTCGAC	1529
QY	3421	GTTGTTGCATGAAGCGGCGGTTCAACCTTCCTGTTTCGTGGGACAGATGACGCCCATTCAGTGC	3480

Db	15292	GTCTTGATGAAAGCGCGCTTCAACCTCTCTGTGTGGGAGAGATCGACGCATCGAGTCG	15351
Qy	3481	CTGTATGATCGTCTGTGCGAGATTTTGTGCTCTGTGAGGCCGATATATGCGCAGCTC	3540
Db	15352	CTGTATGATCGTCTGTGCGAGATTTTGTGCTCTGTGAGGCCGATATATGCGCAGCTC	15411
Qy	3541	CTGATGAGAGCAGGCGCTTCTCGCAATCGGGGTCGTCTCGGTGGGCAATTCAATTTT	3600
Db	15412	CTGATGAGAGCAGGCGCTTCTCGCAATCGGGGTCGTCTCGGTGGGCAATTCAATTTT	15471
Qy	3601	GAACAGATTCCGCGCTTCTCGACATCGCACATTCGAGCTTTACGTACGCGCAGTCGTG	3660
Db	15472	GAACAGATTCCGCGCTTCTCGACATCGCACATTCGAGCTTTACGTACGCGCAGTCGTG	15531
Qy	3661	GGCGGGCGGGTAGACCCCGCGGCAATTCAGATCTGTACGCTCGGTCAAGATTTCTACCG	3720
Db	15532	GGCGGGCGGGTAGACCCCGCGGCAATTCAGATCTGTACGCTCGGTCAAGATTTCTACCG	15591
Qy	3721	AGGCGCGGCAAGACGCGCGCGCCCTCTCCGCGCGGAGCAGCATTGCGGATATGCTT	3780
Db	15592	AGGCGCGGCAAGACGCGCGCGCCCTCTCCGCGCGGAGCAGCATTGCGGATATGCTT	15651
Qy	3781	CGCGACTTCTTAGAGCAAACTAACCGAGTACATGGTGGCTTACGTCTTGTGGAGCTC	3840
Db	15652	CGCGACTTCTTAGAGCAAACTAACCGAGTACATGGTGGCTTACGTCTTGTGGAGCTC	15711
Qy	3841	GATGCGTTGCGCGCTGACGTCAACGCGCAAGTGCATGTAAAGCCCTGCGAGCGGAAG	3900
Db	15712	GATGCGTTGCGCGCTGACGTCAACGCGCAAGTGCATGTAAAGCCCTGCGAGCGGAAG	15771
Qy	3901	GATACCTGTCGCGCGCGGCAATTCGGGCGACACGCGCCACCGGACGCTTGGAGAGATC	3960
Db	15772	GATACCTGTCGCGCGCGGCAATTCGGGCGACACGCGCCACCGGACGCTTGGAGAGATC	15831
Qy	3961	CTGCGCGGCGGTGTAACGGAGGCTCGGGCTGGAAGGTGGTGGGCTTCACAGAGCTTC	4020
Db	15832	CTGCGCGGCGGTGTAACGGAGGCTCGGGCTGGAAGGTGGTGGGCTTCACAGAGCTTC	15891
Qy	4021	GTGCACTTGTGCGCAATCGATTCATCGTTCGTCATGAGAGACCTGTTGACAGAGAG	4080
Db	15892	GTGCACTTGTGCGCAATCGATTCATCGTTCGTCATGAGAGACCTGTTGACAGAGAG	15951
Qy	4081	CTGATAGGAGATGCGCATCACCGAGTTGTTCAAGTACCCGAACCTCGCTCGCTGGCG	4140
Db	15952	CTGATAGGAGATGCGCATCACCGAGTTGTTCAAGTACCCGAACCTCGCTCGCTGGCG	16011
Qy	4141	TCCGCTTTCGCGGAGACTTCGAGAGATCTTAGATCAAGCGCGGCAATGCAAGACGATG	4200
Db	16012	TCCGCTTTCGCGGAGACTTCGAGAGATCTTAGATCAAGCGCGGCAATGCAAGACGATG	16071
Qy	4201	GAGCTTCGCGGAGGAGGAGAGACGTAGCTAA 4223	
Db	16072	GAGCTTCGCGGAGGAGGAGAGACGTAGCTAA 16104	



2257 CTGCGGAGACGACGCGGAAAAAGATCCGCTCATTTGGCAGACAGTTATGGCACACGGT 2316  
2269 AAGTGAAGTGTGAGACTCGGTGCGCGCTGATGCGGATGCTGTCGAGCATTTGAG 2328  
2317 CATGTAAAGCTGTGATGAGAGTGCCTGAGCTGATGAGATGCTGTGG---AATACAC 2373  
2329 GGTGCGCCCGATTCGCTGCTAGTGTCTGCGGCTTTGCTGCTGAGCGGCGACGTGATC 2388  
2374 AGCGCGATTCGAGTATTCGACGTTGCTGCGACCTGTTAGCGGCGACGTGATC 2433  
2389 CCGGTGGGCTGCTGGGCGA---GCTCAGGACATCAGGCGCGGCTGCTGATGATCAGC 2445  
2434 CCGCTAACGTTACCGGACAGATGTCGACGCGCTCAATGAAACGATGACATCAGT 2493  
2446 CTGGCGCGGCGCACCGAAGCGTGCATGCTGCTCATCGGGTACCCCGTGAAGAACGTGAC 2505  
2494 CTGGGTGAGACGACCGAGTGGCCCATCTGTGCTGCTACTACCCGATAGTGAAGTGA 2553  
2506 CTATGTTGGGCGAGCATCCCTTACGCGCGCTGCGCAACGACGTTCCAGCTGCTC 2565  
2554 TCGAGTGGACCGAGTATTCCTTACGCTGCGGCGCTGCGCAACGACGATATACGTGCTA 2613  
2566 GATGAGGCTCTGACACCGCGCCCGCTGCGGCTTCCGCGGCACTTACATTTGGCGGGT 2625  
2614 AATGCGCAACTGAGAGAAATGTCGGTTCGGGTGAGAGAGATTTGATTTGGCGGATG 2673  
2626 GGGCTGCGACTGCGCTACTGCGCGATGAGAGAGACGCGCAAGGCTTCCCTGCTGAC 2685  
2674 GGGCTGCGCAAGGCTTACTGACACGACGAGAAACGCGGCGACGCTTTGTCTGGGCG 2733  
2686 CCGGAGACCGGAGACGCGCTTACAAAGACCGGCGATCTGGGCGCTTACCTGCGCATG 2745  
2734 GAAGGAGTGGTGAAGCAATTTACCGCATGCGGAGATCGCGGCGCTTACTTTGCTGACGG 2793  
2746 AACTGAGATTCATGCGGCGTGAAGACAAACCAATCAAGCTTGGGATACCGGCTTGA 2805  
2794 CAAGTGCCTTTTGGGCGCAACGATCCCAAGTGAAGTGAATGTTACCGTATCGAA 2853  
2806 CTGCGGAGAAATCGAGAAACGCTCAAGTGCATCGCAAGTACGCGGCGGATTTG 2865  
2854 CTGGGGGAAATGAGCGCTGACATTCGCGACATTCGATGTGAGACATGATGTTG 2913  
2866 CCGGTGCGGAAAGACGCGGCGAAACGCTCTTACGCTATGTGATCCCGAGGCGACA 2925  
2914 GCAGTGGGTTA-----TTCTCAACATGTCGG 2940  
2926 CGGAGACCGCGCTGCGGACGAGACCGGACCTCAAGACCGAGCGATGAGCGGAGACA 2985  
2941 CTGTGCTTTTTCGCAACGTCGACGATCGCACGAGCGGCGACATTCAGCTTAAGAA 3000  
2986 CACGCGCGGAGCGGAGCGGCTTGAAGCGACGCGGAGGATGCACTTCAAGCTGCTG 3045  
3001 -----GGGAGGCGGCGGCACTGGCGAGGATTTATTTGTAAATCCGCGACAGCTCTA 3054  
3046 CACGAGCTCCGAGGAGCTTGAACGAAAGCCGCTGCTGATCTGACCGGCGAGATCG 3105  
3055 GCGTTCAAATCTCAAGAGGACATATTCGCGCGCTGATGATGCTGGGATTCGACTGAG 3114  
3106 CGGAGGCGGCGGCTGAGCTCTACGCGGCTGCGCTGATCGGACGTTCTCTTGAAGCC 3165  
3115 GCACCGGCGGATGACACAGTTACATCAACCCCGACGATATCTCATTTTCGCGCGAA 3174  
3166 CCGATTCGTTTGTGATTTGATGATTCCTGAGCTGCTTGAAGAGCGGAGGCGGAC 3225  
3175 AAAACCAAGCTGACAGTGGGGAATTCCTGCTGCGGCTTGGGCAAGATGCTTACCC 3234  
3226 GCGCGACCTTCCCAATTCGTTATCATGCGGCGGACGACATTCGCGGCGGCAAC 3285  
3235 GGTACCTTTTTCGCAAGTATGCTATGCTGCGCGGCGGCTTACCCGCTGCAAC 3294  
3286 TACGCTATGTCAAATCCGCGGATTCGAGGCGCTGAGAGAGGCTTCTATTTATACAC 3345  
3295 TACGTATCTGATTCAGACAAAGATCGAAGGAGATATCGGATATTTACTTAC 3354

3346 CCGTTCAGACCCGTTGCTGAGCT---CTCCGATCAGGAGATGAGCGGAGCGCAC 3402  
3355 CCGGACAGAGCTGTCTTATGCTCGGTAGACACAGAAAGTGAAGTGAAGTGTTCAT 3414  
3403 GTTGGCAAACTTCGACGCTTGTGATGAGAGCGGCTTCAACCTCTCTTGTGGGCG 3462  
3415 GCCGACCTATTCAGCTATTTGCGGATCGGCGGCAATTCAGCTGTTATGCTGAT 3474  
3463 ATGACGCGATTCAGCTGCTGATG---ATGCTGCTGCGAGATTTGCTGCTGAG 3519  
3475 ATGCGGATTCGCTGCTTATGAGGACAGGAGCTTGGACCTTCTGCTGATGAA 3534  
3520 GCCGATATATGAGGCAAGCTCTGATGAGAGCGGCGCTTCTGCAACATCGGCTGCT 3579  
3535 GCAGTACTCTGCGCATTTACTGGAAGAAATGCGCGGCTACGATGAGGCGGTG 3594  
3580 CCGGTGGGCAATTCATTTGAAACAGGTTGCGCGGCTTCTGACCTGCAATTCGAC 3639  
3595 CAATGGGATGAGGACCTTTCCGCTGCTGATGCAATTTTCAATGCTCCGACATGAT 3654  
3640 GTTACGTCAGCGGATGCTGGGCGGCGGTAGACCGCGGCGAGTTCAGGTCTGATG 3699  
3655 CCGTATGCTATGACCGGTGCGGCGGCGCATGAGGCGGCAAGTGTGACACA 3714  
3700 CTGCTGAGATTCCTCAACGAGGCGGCGCACAGCGCGGCGGCTTCCGCGCGGAG 3759  
3715 TTGCGGCGGATTTCTCAGCTATGAGAAACGAAAGAAACGCTGCGCGCTGAGCATG 3774  
3760 CAGCATTCGCGGATATGCTTGGGACTTCTTGAAGACAAATCTACCGAGTACGTTG 3819  
3775 CAGAGCTTACAAAGATGCTGCTGCGGCGCTGCTGAGCAATCTGCTGATATATG 3834  
3820 CCTACGCTTCGAGAGCTGATGCTGCGGCTGAGCGCTGCAACGCGCAAGTGTGATG 3879  
3835 CCGATGATCTGATGTTAGCAGCGATTTCTGTTAACGCTTAAACGCGACCTGATG 3894  
3880 AAGCGCTGCGGAGCGGAGATTAATCTGTCGCGCGGCAAT-----TCGGGCGACG 3933  
3895 CAATAATTCAGCTGCGAGGCGGAAATAATTCACGAGCTGACGCGGTGCTCAATC 3954  
3934 GCGCGACGAGCGCTTGGAGAGATCTCTGCGGCTGCTGACGAGGCTGCTGCGCTG 3993  
3955 CAGGTGACAGTGTGCTTAAACAGCGGCTGCTGCGCTGCTGCGAGGATCTGCGCTG 4014  
3994 GAGTGTGCGGCGCTCAGACAGCTTGTGATCTTGTGCGATGATGATCAGATGCT 4053  
4015 AGCCAGCTGCTGCGGAGAGATTTCTGCTGCGGCGGCGAGTCTTATGAATG 4074  
4054 CGCATGAGAGCTGTTGCAAGAGGCTGATGAGAGATCGCATCAGGATGTTTC 4113  
4075 CGTATTCAGAGGACCTTGAAGGCGATTAATCGGACAGAGATTCATGTTGATGCTTC 4134  
4114 CAGTACCGAACCTCGGCTGCTGCGG 4140  
4135 GGTGCGCAACCATCGCGGATGTCGCG 4161

RESULT 3  
US-10-238-075-921/c  
; Sequence 921, Application US/10238075  
; Publication No. US20030148324A1  
; GENERAL INFORMATION:  
; APPLICANT: I. N. S. E. R. M.  
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat  
; FILE REFERENCE: BLANDINE  
; CURRENT FILING DATE: 2002-09-10  
; PRIOR APPLICATION NUMBER: 0003145  
; NUMBER OF SEQ ID NOS: 1576  
; SOFTWARE: Patent in version 3.1

SEQ ID NO 921  
LENGTH: 48715  
TYPE: DNA  
ORGANISM: Escherichia coli  
US-10-238-075-921

Query Match 12.8%; Score 543.4; DB 12; Length 48715;  
Best Local Similarity 48.5%; Pred. No. 1.1e-131;  
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;

181 ATCTGTCCTCCGCGCCGAGCGGCGGCGCTGCTCTCTCAAGACATCCAGATCC 240  
18373 ATAGTGACAGACCTGAGCATGTATCAACGGTTTCCATGTAGCGATTTAGGCGCC 18314  
241 TACTGGCTGGGCGCGAGAGCGTTTACGTCTCCAGCGGATCCAGCCTATGCGGA 300  
18313 TACTGGCTTGGAGCG--ACAGACGGGGTGTACTCTCATGCGCACCCATCTTACATGAA 18257  
301 TACGACTGTAGGATCTCGACGTGCGGAGCTGAGCGGCGCTTTCGAAAGTCGTGCG 360  
18256 TTTGACGTAGAACACTTTATATGTATGCGCTTTTACCATGCGGTGAATCGCTGATCGT 18197  
361 CGGCAAGCATGCTTTCGGGCGCCACAGCTGCGCGACATGATGAGGTGATCGACTTAA 420  
18196 CGCCATGAAATGCTATGCTGCGGGGTACTCCCGACGTAATCAGAGATTTCTGGCGCA 18137  
421 GTTCAGCG--CGACATCGAATCATGATCTGCGGGGCTCGACCGGAGCACAGGGAA 477  
18136 GTGCGCGCTATCATGTTAGTAGAGACGCGATCTGAGTCTTGTCCCTTAACGACGAAAC 18077  
478 GGGAGCGCTGCTGCTGTTGGAGATGAGATGCGACCGCATGTATGACACGAGCGCCT 537  
18076 GATGCTTATGAGCGATCCGCGATCGGCTGTGCAATGTGATCCGCGAGTCTTTGG 18017  
538 CGGCTCTATCAGCTGCTGCGCGCTGCGCTGAGCAGACGGCGAAACCGCTCTGCTCACT 597  
18016 CGGCTGTTGATTTGATGATTATTCGCTTGCACGGGCGCAATGGCGCGCTTGATTTCACT 17957  
598 ATGATCTCTATTACGTTACCTTAGGACGCTGTCCATCATCTTCAAGGATGCTCAGC 657  
17956 CTCGATCTGCTCATTTGCTCCATGCTCTGATGTAGCGACGCTACAGCAGAGTTGATGAG 17897  
658 TTCTCGAAGATCCCGAGACCTCTCTCCGTCCTGAGCTCTGTAACCGCATTTATGTA 717  
17896 CTGTACCTGAGGCCCATGTGTACTGCGGTGCTACCGTTCTCTTTTGTGACTAGTG 17837  
718 CTCGCGCTGAGTCTCGCAAGAAAGTCTGAGCGCATCAAGATCATGATGATTACTGAG 777  
17836 CAGGCGCTGTTGGTAGAGCAGGGGAGTAGAAGCTATGCAAGCGATCAGGCGCTATGGCA 17777  
778 CGGCGCATGCGGAGCTCCCACTCCGCGAGCGCTTCCGATGAGGCGCATCTTACC 837  
17776 CGGCGCTGCGCGACGCTGTATGCGCCACCAACGCTGCCGTACAGGCGCATTTGGCGCA 17717  
838 CTGGAAGGATCGGCTTCCGGGACACAGAGCAATGCTGCGGAGCTCTGCGGGTGA 897  
17716 CTGTCTGCAATCAGTTGCTGATGCTGCCGCTCATCGCTGTCAAGCCACAACTGGGAGTG 17657  
898 TTGAAGCGGCGTGTGCGGAGCGCGGCTGACCCCGACGAGCGCTCATCTGTGCTGATTT 957  
17656 CTGAGGCGCTGCGCCACAGCAACAGTATCACCAAGACGCAATGTTGTTGACAGTCTTT 17597  
958 TCCGAGGTATGCGGCGCTGAGAGCGGAGCGCGCTTTCAGCTCAACATTAAGCTCTTC 1017  
17596 ACGCAAGTCTGCGCAGTTGAGGCTTACCGCGACCTTTACGCTCATCTGACGTGTTC 17537  
1018 AACCGGCTCCCGTCACTCGGCGGTAAGATATACCGGGGACTTCACTGATGATGTC 1077  
17536 AACCGCGCGAGGTTTACCCCAACGAGAGCGATTTATGTTGATTTTACCGCTGACG 17477  
1078 CTCCTGAGATCGACCACTCGCGCAAGAGCTTGGAAACAGCGCGCTATGACTTCA 1137  
17476 TTGCTGAATGTTGTTAGCAGCGCAGCACTTTATGCGCCACAAAGCTCAGGCTATTCAG 17417

1138 GAGCAGCTGTGGGAAGCATGATGATCTGCGACGTAAAGCGGAT-----CGAGGTC 1188  
17416 GTGCAACTGTGGGAAGATCTGAAACATGCTGCTTTCATGTGGATCGCGCCAGCGAGCG 17357  
1189 CAGCAGAGAGCGCGCGCGGCTCTGGGAGATCCAAACAGAGCGGATTTGTCCTCGGTGCTC 1248  
17356 CTGATCCATAGCGGTGCTTTCATATGCGCGAGATGCGCGGTGATTTACATGATGTTGAT 17297  
1249 ACGAGCGGCTTAAACAGCAAGTGTGTGTACCTTGTTCAGAGGCTCGGAATCCG 1308  
17296 ATCGAGGGGAGAACACTGCGCAAGACCTCGGGGACAAACCGCTTTTACTGTGTGCTG 17237  
1309 GTGTACACAGACCGCACTCTCTAGCTGCTGCTGATCAATCACTGATGAGACGAT 1368  
17236 GACGCCAATATTATCCCAAAACCGAGGTGTGCTGTGATCAACAGATATGAGTTGGCT 17177  
1369 GGGAGACTGTCTCTGCGTGGGACATCGTCAACGAGAGTTTCCCGCCGACCTTGTGAC 1428  
17176 GGGAGTTGCAATTTCACTGGGACCGGTCGAGCACTGTTGATATCACAGCTGTGGAT 17117  
1429 GACATGCTGAAAGCTACGTGTTTTTCTCGGCGGCTCACTGAGAAACATGGGTGA 1488  
17116 CAGATGTTGTGCTTATTTGATATGCTGACAGCGGCTGTGTCATGCC-----G 17066  
1489 CAGGTGCGCTGTTCCTCCGCTGCCAGCTAAGAAAGCGGGCGAGCGCAAAACGCGACC 1548  
17065 CAAAGTTGTGGGGGTAAATGATGTTCTGTGGCGCTCCACCGCTTATGTGACCGGTACG 17006  
1549 AACGCGCTGTAGAGGAGCATATGCTGACAGGCGCTGTTGCGGCGCGGCTGAGCAGTGT 1608  
17005 CAGGCTCTGACACTACCGCTTGTGTGACATGATATTAAGTGTGATGAGCAGCACTGAC 16946  
1609 CCATGACGCTGCGGTGTGTGCGCGCAAGACGCTCAAGTACGAAGCTTTTGGCGC 1668  
16945 CCACAGAAACTGCGCTGATAGTCTTATCCGTGAATGACCTATGCGCAACTGTGACG 16886  
1669 CGTTGCGCGGCACTTGGGCGCGCGGTGCGGAGCAGGGGGGACCGCCGCAACATTTGTC 1728  
16885 GCGGCGAATCATGTGCGCGCGCTGTTGATGCGCTGAGGCGTGCAGATGCGACCGCGTG 16826  
1729 GCGGTGTGATGAGAAAGCTGAGGAGGATGTGCGGCTTTCGCGGTCTGAGTCA 1788  
16825 GCGGTGTGATGAGAAAGCTGAGGAGGATGTGCGGCTTTCGCGGTCTGAGTCA 1788  
1789 GCGGCGCTTACGTGCGCATGATGATCGACCTTCCGCGAGCGTATCACTACTCTC 1848  
16765 GGTGCGGTCTATCTGCGAGTGTATCGGTGCTAACCGCACAGCGTTCGCTTTGCTG 16706  
1849 GATCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1908  
16705 ACGGTGGGAGGTGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 16646  
1909 CCGCGGGGATCCAGCGGCTGCTGTGAGCGAGCGCGGCTGAGAGCGAGCGGACGAG 1968  
16645 CTGCGCG-----TGTGATCATGCAAGCGAAGATGCTGAGACCGCTGCTGCG 16598  
1969 CCTCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2028  
16597 CGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 16538  
2029 TCCAGAGGTTGCCAAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2088  
16537 TCCACCGGATCCCGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 16478  
2089 GACATCAAGAGCGCTTGAATATGAGGCGCGGAGCAGAGTGTGCTGCTCTCTGCTG 2148  
16477 GACATCAAGAGCGCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 16418  
2149 ACCTTGATCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2208  
16417 ACCTTGAACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 16358

2209 GTGCCGACCGCTCCAGCTGCGCATTCGCGCATTTGGCAGAGTTGATCGAAGAG 2268  
Db CTGCCGAGACGAGACGGAAAAAGATCCGCGTCACTGGCAGACAGTTATGGCACACGCT 16298  
2259 AAGGTGAGGTGTGAACTCGGTGCGCGCTGATGCGGATGCTGTGAGCATTTTGA 2328  
Db CATGTAACGCTGTGAAATGCAAGTGGCCCACTGATGACAGATGCTGTGG---AATACAC 16241  
2329 GATGCCCGCCGATTCGCTGCTAGTCTCTGCGGCTTTGCGTGTGAGGCGGACATGATC 2388  
Db AGCGCGCATCGGATGATGATTAATCCGACGTTGCGTCTGGCACTGTTAAGCGGACCTGATC 16181  
2389 CCGGTGGGCTGCTGGCGA---GCTCCAGGCGCATCAGCCCGCGGTGTGGTATCAGC 2445  
Db CCGCTAATGTTATCCGAGCAGATGCGCAGCGGCTCAATGAAACATGAGACATCATAGT 16121  
2446 CTGGCGGGGCGCACCGAAGCGTTCATCTGTTCATTCGGGTACCCGCTGAGAGACGTGCAC 2505  
Db CTGGGTGAGGCGACCGAGTGGCGCATCTGTGCTACTACCCGATAGTGAAGGTGAA 16061  
2506 CTATCGTGGGCGAGCATCCCTTACGCGCGCTGCGCAACAGACGTTCCAGTGTCTC 2565  
Db TCGAGTGGACCAAGATTCCTTACGCTGGGCGCTGCGCACACGATATACGCTTA 16001  
2566 GATGAGGCGCTGAAACCGCGCCGCTGTGGGTTCCGGGCGCACTTACATTTGGCGGGTC 2625  
Db AATGGCAATCGAAGAAATGTCGGGTCCGGGTGAAAGAGATTTGATTTGGCGGGATG 15941  
2626 GGGCTGCACTGGGTTACTTGGCGCATGAAAGAAAGCGCGAAGCTTCTGTCGAC 2685  
Db GGGCTGGGCAAGGCTTACCTGAAACGACAGAAACGCGCGCAAGCTTTGTCTGGGCGC 15881  
2686 CCCGAGACCGGGAGCGCTCTCAACAAGACCGGCGATCTGGGCGGCTTACCTGCGGATGA 2745  
Db GAGCGATGCTGAGCAATTTACCGCATGCGGATCCGCGGCTTACTTGTGACGAG 15821  
2746 AACATCGATTATGCGGCGTGAAGCAACCAATCAAGCTTCCGGAATCCGCTTGA 2805  
Db CAAGTCGCTTTTGGGCGCAAGATCCAGTGAAGTGAATGTTACGTTACGAA 15761  
2806 CTGGGGAATTCGAGAAACGCTCAAGTCGATCCGAACGTAACGCGGCTGATTGTG 2865  
Db CTGGGGAATTCGAGAAACGCTGATTCGCGACATCCGATGTGAGCACTGATGTGTG 15701  
2866 CCCGTCGGGAAACGACGCGCGCAACAAAGCTCTTACGCTATGTTGTCGCCGAGGCGCA 2925  
Db GCAATGGGTA-----TTCTCAACATGTGCG 15674  
2926 CGGAGACCGGCTGCGGACAGACGCGACTCAAGACCGAGCGATCGACGCGAGCA 2985  
Db CTGGTCGCTTTTGCCTCAATCGATGCGCACGAGGCGAGGCTTGCAGCTAAGAA 15614  
2986 CACGCGCGCGAAGCGGCTTGAAGCAAGCGGAGGCTGAGTTCAAGTCTGCTGA 3045  
Db -----GGGAGGCGCGCGCACTGGCGGATATATATGTAATCCGCAACAGGCTGA 15560  
3046 CACGCACTCCGAGGAGCTTGAAGCAAGCAAGCCGCTGCTGATCTGAACGCGAGATCGG 3105  
Db GCGTTCAATCTCAAGAGCACATATTCGCGGCTGATGATGCTGGGCAATTCGACTGACG 15500  
3106 CGGAGGCGGGGCTGAGCTTACGCGGCTGCGCTAGCGCTCGAAGCTTCTTGAAGCC 3165  
Db GACCGGGGAGATGACACAGTTACATCAACCGCGCAGCTATTCGATTCAGCGCGAA 15440  
3166 CCGATTCCGTTTGTGATTTGCTGATCTGAGCTGCTTGAAGCAGCGTGAAGCCGAC 3225  
Db AAAACCAAGCTGCGCACAGTTGGGGAATTTGCTGTGGGCTTGGGCGCAATGCTTACCC 15380  
3226 GCGCGGACCTTCCCAAAATTCGCTATCATCGGCGGCGACAGCTACCGGTGAAC 3285  
Db GGTCTACTTTTGGCAAGTATGCTATGCTGCGCGGGGCGCTATACCGGTGAAC 15320  
3286 TACGCTATGTCAAATCCGCGCGCATCGAGGCGTGAAGGCGCTTCTATTTATACAC 3345

15319 TACGTATACCTGATCCAGACAGATCGAAGAGGAGATCCGGTATTTACTTCTGAC 15260  
3346 CCGTTCCAGACCGCTTTGCTGAAGCT---CTCCGATCAACGGGATCGAGCGGAGCGAC 3402  
Db CCGGACAGAGCTGTCTTATGCCGTAGACACCAAGTGAAGTGAAGAGTGGTTTCTAT 15200  
3403 GTTCGGCAAACTTCGAGCTGTGATGAAGCGCGCTTCAACTCTGTTGCGGCGAG 3462  
Db GCCGACCTTATCACTATTTCCGATTCGCGGCGGATTCAGGCTTTATATGTTGCTGAT 15140  
3463 ATGACCGCATGATGCTGCTGATG---ATGCTGTGCGAAGATTTTGCCTGCGAG 3519  
Db ATGCGGATGATCTGCGCATTTATGCGCAAGAGCAGCTTGGCACTTCTCGGTATGAA 15080  
3520 GCCGATATATGCGGAGCTCTGATGAGAGCGCGCTTCTGCAACATGAGGCTGCT 3579  
Db GCAAGTACTCTGCTCCATTTACTGAAAGAGATGCGCGCTACGATGAGTTGGGCTGTG 15020  
3580 CCGGTGGGCAATTCATTTTGAACAGTTGCGCGGCTTCTGACCTGCAATTCGAGC 3639  
Db CAATTTGGATGAGGACATTTTCCGCTGTGCGCATGCAATTTTCAATTTGCGCACATCAT 14960  
3640 GTTATGTCACGCGATGCTGGCGCGGCTGACCCGCGCAATTCAGGCTCTGACG 3699  
Db CGTATGTCATTCGACCGTGGGCGCGCATGAGGCGCAAGGCGGCAAGTGTGACACA 14900  
3700 CTGCGTCAAGATTCCTCAACGAGGCGCGCACAGCGCGGCGCCCTCCGCGCGCAG 3759  
Db TTGCTGCGCGATTTCTCACTTATGAGAAACCGAAGAAACCGCTGCGCGCTGACATG 14840  
3760 CAGCACTTCGCGGATTCCTTCGCGACTTCTTGAAGCACTTACCGGATTAATGCTG 3819  
Db CAGAGCTTCAAAAGATGCTATGCTGCGGCTGCTGCTGACGAACTGCTGATATATGCTG 14780  
3820 CTTACAGTCTTGTGAGGCTGATGCTGCTGCTGACGTCACACGCGAGCTGATGCT 3879  
Db CCGAGTATGATGATGTTAGCAGACGATTTTCGTTAACCGCTAACCGGCACTGATGCTG 14720  
3880 AAGCGCTCGCGGACCGAAGATATCTCTGCTGCGCGGCT-----TCGGGCGACAG 3933  
Db CAAAAATTAACGCTGACGAGGCGAAATTTGCCACACGCTGACGCGGTGGTCAATC 14660  
3934 GCGCGACGAGGCGCTTGAAGAGATCTCTGCGGCTGCTGATGAGGAGGCTGCGGCTG 3993  
Db AGCAGGCTGCGGCGGAAAGATTTCTTCTGCTGCGGCGGCGAGTTCTATAGAAATGCTG 14540  
4054 CCGATGAGAGCTGTTGCAAGAGGCTGATGAGAGATTCGCAATCCGAGTGTTC 4113  
Db CGTATTAAGCAGGACATTTGAGGCGATTAATGCGGAGAGATTTCCATTTGATCTGTTG 14480  
4114 CAGTACCGAATCTCGGCTGCTGCGG 4140  
Db CGTCTGCCAACATTCGCGGATGTGCG 14453

RESULT 4  
US-10-085-959-57/c  
; Sequence 57, Application US/10085959  
; Publication No. US20030165870A1  
; GENERAL INFORMATION:  
; APPLICANT: Blatner, Frederick R.  
; APPLICANT: Welch, Rodney A.  
; APPLICANT: Burland, Valerie D.  
; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073  
; FILE REFERENCE: 960296.97648  
; CURRENT APPLICATION NUMBER: US/10/085, 959  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: 60/242, 412



```

; PRIOR FILING DATE: 2000-10-
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version
; SEQ ID NO 57
; LENGTH: 50538
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-085-959-57

```

Query Match	12.8%	Score 543.4	DB 12	Length 50538
Best Local Similarity	48.5%	Pred. No. 1.1e-131		
Matches 1933; Conservative	0	Mismatches 1961	Indels 93	Gaps 12

OY	181	ATCGTGGCCGCGCCGAGCGCGGACAGCTCGGTTTCTCTCCACAACTCCAAAGATATCC	240
Db	20201	ATATGTGACGACCTTGGATCGATCTCAACCGTTCCATTGACGGATGTTACAGCGCGCC	20142
OY	241	TACTGGCTGGGCGCGGACGAGAGCGTTTACGGTCCCAAGCGGGATCCAGCTATCCGCGAA	300
Db	20141	TACTGGCTTGGAGCG--ACAGACGGGGTGTTAACTCGATCGCGACCACACTTACATGAA	20085
OY	301	TACGACTGTACGAGATCTTCGACGTGCCGAGGCTGAGCCGCGCCTTTGGGAAAGTCGTCCG	360
Db	20084	TTTACACGTAGAAACCTTTAATGTTCACCGGTTTTTACCATGCGGTGAAATGCGCTGATCGCT	20203
OY	361	CGGACAGACATGCTCTTGGGGCCCAACGCGTCCCGGACATGATGAGGTGATTCGAGCTTAA	420
Db	20024	CGCATGAAATGTCTACGTGGGGCGGGATCTCCCGGAGGTATCTGAGAGAAATTGCGGCGAA	19965
OY	421	GTCGACGC--CGACATCCGAGATCATGATATCTGCGGGGCTCGAACCGGAGACACGGGAA	477
Db	19964	GTGCGCGCGGTATCTAGTTAGAGACAGCGGAGTCTGATGTGTGTCCTTAAACGACCAAAAC	19905
OY	478	GCGAGGCTCGGTGTGTTGACGAGATGCGATGTCCACCGCATTTATGACACCGAGCGCCT	537
Db	19904	GATGCTTGTATGGGATCTCCGCGATCGGCTGTGCTGATCATGTGATATCCGCAATGCTTGG	19845
OY	538	CGGCTTATATCAGTGTGCGCCGTTCCGCTGGAGAGAGCGGAAACCCGTCTGTGCTCACT	597
Db	19844	CCGCTGTTGTTGATTTCAGTTATTCGGGCTTGGACGGCGCAATGAGCGGCTTGACATTCAGT	19785
OY	598	ATGCACTCATTTAAGTTGACCTTAGAGAGCGCTGTCCATCATCTTTCAAAGACGCGCTCAAC	657
Db	19784	CTGCATCTGCTGATTTGCCAGTACTCTGAGTATGGGACGCTACAGCAGGAGTTGATGATG	19725
OY	658	TTTCTACGAAGATCCCGAGACCTCTCTCTCCCTGTGAGGCTCTGTATCCGCGATTTATGTA	717
Db	19724	CTGTACCGTGAAGCCCAATGTGTCACTGCGGTGCTACCGGTTCTTTTGTGTACTACGTG	19655
OY	718	CTGCGGCTGTGAGTCTCGCAAGAAAGTCTGAGCGGCTCAACGATCATGATGATTACTGGAA	777
Db	19664	CAGGCGCTGTTGTAGAGCAGGCGAGGAAAGCCTATGACAGCGCATCAGGCTATTATGGCAA	19605
OY	778	CGGCGCATTCGCGCAGACTCCACACTCCGCGCAGCTTCCATGTAGAGGCCGATCATCTACC	837
Db	19604	CGGGCGGCTGCCGACAGCTGTATGTGGCCACCAACGCTGCGCTAAGGGCGATTTGGCGCAA	19545
OY	838	CTGAAGAAGATCCGCTTCCGCGACACAGGAGCAATGCTGCGGTGGGACTCTCTGGGGTCCA	897
Db	19544	CTGTCTGTGAGTCAAGATTGTGATCGTGCCTGTATCTGCTGTGCAACCCCAAACTGGGGGAGTG	19485
OY	898	TTTGAAGGCGCGGTGTGCGGGGAGCGCGGCGGTACACCCGACGGGGCGTATCCGCGGTGATTT	957
Db	19484	CTGAGCGGCTGTGGCCCAACGACACGATCATCCAAAGCGGCAATGTTGTGACAGTCTTT	19423
OY	958	TCCGAGGTGATCGGGCGCTGTGAGCGCAGAGCCCGCGTTTACGCTCAACATAACGCTCTTC	1017
Db	19424	AGCCAAAGTGTGGACAGCTGTGAGGCTTAAAGCCGAGTTTACGCTCAATGTGAGTTGTTCC	19365
OY	1018	AAACCGGCTCCCGGCTCATCTCGCGGTGAACGATATCACCGGGGACTTCAAGTGTGATGTC	1077
Db	19364	AAACGCGCGCAGGGTTATCCCAACGACAGGACAGTTATGTGTATTTTAAACCGTGTGACG	19305

OY	1078	CTCTGGACATTCGACACACCACTGGCGGCAAGAGCTTGGAAACAGGGCGCTAAGGATTTCAA	1137
Db	19304	TTGCTGAATGTTTGTTAACGACAGCCAGCACTTTATGCCCAACAGCTCAGGATATTCA	19245
OY	1138	GAGCAGCTGTGGGAGAGCGATGATCTACTGCGAGCTAAGCGGAT-----CGAGTTC	1188
Db	19244	GTCGACCTGTGGGAAGATCTCGAACTACGTCGTTCACTGAGGGATTCGGCCAGCGAGGGC	19185
OY	1189	CAGCGAAGAGCCGCCGGGTCTCTGGGGATTCGAACGAGGGCAATTGTTCCCGGTGCTC	1248
Db	19184	CTGATTCATAGCCGTCCTTTTCATAGCGCCGATCCGAGTGGGATTCCTAGTATGTTGGAT	19123
OY	1249	ACGAGCCGCGCTTAAACGACGAAGTCGTGTGTCACCTCGTTGCGAGAGCGCTCGGAATCCG	1308
Db	19124	ATCGACGGGGAGACGATCGGCGCAAGACCTCGGGAGACAAACCCGTTTACTGTGTCCG	19065
OY	1309	GTTGTAACACGACACGACGACTCCTCAGCTGTCTGTGATCATGCTTACGACAGAT	1368
Db	19064	GACCCCAATTTATCCCAACACCGCAGGTGTGCTGATCACCAGGTGATCGAGTTGGCT	19005
OY	1369	GGGGAACCTCGTCCTCGCGGTGGGACATCGTGAACGAGAGTGTCCCGCCGACCTTGTGAC	1428
Db	19004	GGGAGTTTCGATTTCAACTGGACGGGGTGAACACTGTTTGATACACGCTGTGGAT	18944
OY	1429	GACATGCTCGAAGCGTACGTGTTTTTCTCCGCGGCTCACTGAGGAACCATGGGGTGA	1488
Db	18944	CAGATGTTTGTGCTTATTTGTCAATGCGCTGCAGAGCCCTGATGGCCATAGCC-----G	18894
OY	1489	CAGGTGGCGTTCCTCGTCCGCCCTCGCCAGCTAAGAACGGCGCGAGCCCAACCGGACC	1548
Db	18893	CAAAATTGTGGGGGGTAAATAGTTCTTGCGCGCTGCCACCGTTATGTGACCGGTACAG	18833
OY	1549	AACCGCTGCTGACGACGACATPACGCTGACAGGCTTGTGACGGCGGGTGCAGCACGTC	1608
Db	18833	CAGGCTCTCGACCTACAGGCTTGTGTGACCATGAGATPACTGGTCAAGGACGACTGAGC	18774
OY	1609	CCCATGCAGCTCGCCGTGTGTGTGGCGCGCAAGACGCTACGTAACGAAGCTTTGCGCC	1668
Db	18773	CCACAGGAACCTGGCGTGAATCAAGTCTTATCCGTAAATTGACCTATGCGCAACGTCTGAGC	18714
OY	1669	CGTTTCGGGCGACTTGGACGCGCGGCTGCGGACGACAGGGGCAAGCCGCAACACTTGTGC	1728
Db	18713	GCGCGGATATATGTGCGCCGCCCTCTTATAGCGCTGGGCTGAGATAGGCAACCGCGTG	18655
OY	1729	GCGGTGTGTATGAGAAAGGCTGSGAGCAGGTGTGCGGTCTCGCGGTCTCGATCA	1788
Db	18653	GCGGTGTGTATGAAAAAGGCTGGCAGCAGATTTGCCGCGTACCGGCAATTTTACACATG	18594
OY	1789	GGGCGGCGCTAACGTCGCGGATCGATGCGCGACCTPACCGGCGGAGGTATCCACTACCTCTC	1848
Db	18593	GGTGCGGTCTATCTGACAGTGAATCCGATGCTAACCGCACAGGCTGCCAGCTTTGTGTC	18533
OY	1849	GATCATGTGATGAGTAAAGCTCGTGTGACGACGACGACCATGCTGTGATGGAACCTGTGATG	1908
Db	18533	ACGGTGGCGGAGGTGCGGGTAAAGTAAACGACACCGGATCTCACGAATTGAGACGCTG	18474
OY	1909	CCGCGGGGAGATCCAGCGCGCTGCTGTGAGGAGAGCGCGCTCGAAGGCAAGCGGACACAG	1968
Db	18473	CTGCCCC-----TGTGATATCATGACGAGAAAGCTGACACGCTGCTGG	18422
OY	1969	CCTCCGATGTATGCCCATTTACAGACCTTGGATCTCGCGTATGCTATCAACTTCGAGGA	2028
Db	18425	CCGTTGCTTAAGTGTGGGGAGTGTACGGAATCTGGCCATATATTTTCACTTCGCGG	18366
OY	2029	TTCACAGGTTTGCCCAAGGGGGTGAATGATCATGTGGGGTCCGTCAACACATCTGTG	2088
Db	18365	TTCACCGGTTACCCCGAAGAGTGAATGATGACACACGTGCGGCAATGAACAGCTGGAA	18306
OY	2089	GACATCAACGAGCGCTTCAAAATAGGCGCCGAGACAGGGGTGTGCGCTCTCTCGCTG	2148
Db	18305	GACATCAACGAGCGCTTTGGCTCAATGCGCAGGATAGGGGTGTCCGGCTGTATCATTTG	18244
OY	2149	AGCTTCGATCTCTGATCTATGATGTGTTCCGGATCTGTGCGGCGGCGGTATGATCTGTG	2208



18245 AGCTTTGACCTGCTGCTTACGATGCTTTGGCCCTTTTATGTTGGTGCAAGCCGCTGCTA 18186  
18245 GTGCGGAGCGGCTCAAGCTGGCGATCGGCGCATTTGGGCAAGATTTAGTGAACAGAG 2268  
18185 CTGCGGAGAGAGAGCGGAAAAAGATCGGCTCATTTGGCAGACAGTTATGGCAGACGCT 18126  
2269 AAGGTGACGCTGTGMACTCGGTGCGGCGCTGATGCGAGATGCTGTCGAGCATTTTGG 2328  
18125 CATGTAAAGCTGTGAATGCAAGTGCCTGACATGATGCAAGATGCTGTGG---AATACAC 18069  
2329 GGTGCCCCGATTCGCTGCTAGTCTGTGGGCTTTGGCTGTGAGCGGCGCATGTGATC 2388  
18068 AGCGCGATCGGATGATTAATCCGACGTGCGCTGCGCATGTTAGCGGCGCATGTGATC 18009  
2389 CCGGTGGGCGCTGCTGGGCGA---GCTCAAGGCAATCAGGCGCGGCTGTGATGATGAC 2445  
18008 CCGCTAACCTTACCGGAGCAAGTGCAGCGGCTCAATGAACGATGACATCATCATGAT 17949  
2446 CTGGCGGCGGCGCACCGAAGCGTGCATCTGATCATCGGATACCCGCTGAGGAACTGTGAC 2505  
17948 CTGGGTGAGAGGAGCCGATGCGCCATCTGTGCTGCTACTACCGCATAGTGAAGTGA 17889  
2506 CTATCTGGGCGAGCATTCCTTACCGGCGCTGCTGCGCAACAGACGTTCCAGCTGCTC 2565  
17888 TCGACGTGAGACAGATATTCCTTACCGGCGGCTGCGCAACAGCATGATATGCTGCTA 17829  
2566 GATGAGGCGCTGMAACCGCGCGCGCTGCTGCTGCTGCGGCGCACTTACATTTGGCGGCTC 2625  
17828 AATGGCAACTGTGAGAAATGCTCGGTGCGGCTGAGAAAGATTTGATTTGGCGGAGT 17769  
2626 GGGCTGCACTGAGCTACTGCGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2685  
17768 GGGCTGCGCAAGGCTTACTTGAACGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 17709  
2686 CCGGAGACCGGAGAGCGCTTCAAGACCGGCGATCTGGGCGCTGCTGCTGCGGAGTGA 2745  
17708 GAAGGAGATGATGAGCAATTTACCGGCTGCGGAGATGCGGCGCTGCTGCTGAGCGG 17649  
2746 AACTGAGATTCATGGGCGCTGAGAGACCAACCAATCAAGCTTGGGAGTATCGGCTGAG 2805  
17648 CAAGTGGCTTTTGGGCGGCAACGATCCCAAGGAGGATGATGATTTACCTGATTCGAA 17589  
2806 CTGCGGAGAAATCGAGAGAAAGCTCAAGTGCATCCGAAGTACGCGGCTGATTTG 2865  
17588 CTGGGAGAAATCGAGCGCTGATTTGCGGACATCCGATGTTGAGACATGATGATGATG 17529  
2866 CCGGTGCGGAAAGACGCGGCGAACAAGCTCTTACGCTATGTTGTCGCGAGGAGCA 2925  
17528 GCAGTGGGTA-----TTCTCAACATGTCGG 17502  
2926 CCGAGACCGCTGCGGAGACAGACCGGAGCTCAAGACCGAGCGGATGAGCGGAGAGCA 2985  
17501 CTGTGCTTTTTCGCAACGATGACATGCGACCGGCGAGGCAATTCAGGCTTAAGGAA 17442  
2986 CACGCGCGGAGAGCGGAGCTTGAAGCGGCGAGGAGTGCAGTTCAAGTGTGCTGCA 3045  
17441 -----GCGAGGCGGCGGAGCTGCGAGGATTTATTTGATTCGCGGACAGGCTTGA 17388  
3046 CACGAGCTCGGAGGAGCTGAGCGGAAAGCGGCTGCTGATCTGACCGGAGGAGATCG 3105  
17387 GCGTTCAACTCAAGAGAGCAATATTCGCGGCTGATGATGCTGAGGATGCACTGAC 17328  
3106 CCGGAGGCGGCGCTGAGCTCAAGCGGCTGCGGCTGCTGCAAGCTTCTTGAAGCC 3165  
17327 GCACCGGCGGATGACAGCTTACATCAACGCGGAGCTATGCTATTTACGCGGCA 17268  
3166 CCGATTCGTTTGTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 3225  
17267 AAAACCAAGCTGAGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 17208  
3226 GCGCGACCTTCCCAATTCGTTATCATGCGGCGAGACGATACCGGCTGCAAC 3285

17207 GGTCTACTTTTGGCAAGTATGCTATGCTCCGCGGCGGAGGCTATACCGGCTGCAAC 17148  
3286 TACGCTATGTCATATTCGAGCGGCGGATGAGAGGAGCTGAGACGAGGCTTATATTAACAC 3345  
17147 TACGCTATGTCATATTCGAGCGGCGGATGAGAGGAGGATGAGGATTTACTTCTGAC 17088  
3346 CCGTTGAGCAGCGTTTCTGAGACT---CTCCATACCGGATGAGCGGAGCGGAC 3402  
17087 CCGGAGAGAGCTGCTTATGCTCCGATGAGCAGAGGCTGAGTGAACAGTGTTCAT 17028  
3403 GTTGGGCAAACTTCGAGCTTGTGATGAGAGGCGGCTTCAACCTCTCTTGTGGGACG 3462  
17027 GCCGAGCTTATCAGTCTATTTGCGGATCGGCGGCTTACAGCTGTTATGATGCTGAT 16968  
3463 ATGACGCGCATGAGTGTGCTGATG---ATGCTGTGCGAGAAATTTGCTGCTGAG 3519  
16967 ATGCGGCTGATCTGCTCATTTTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16908  
3520 GCCGAGATATGAGGAGAGCTTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3579  
16907 GCAGGATCTCTGCTCATTTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16848  
3580 CCGGTGGGCAATTCATTTTGAACAGGTTGCGGCTTCTGACCTGCGCATTTGAGC 3639  
16847 CAATTTGGAGTGGAGAGCTTTTCCGCTGTGCTATGATTTCAATTTGCTGCGCATCAT 16788  
3640 GTTACGTCAGCGGATGCTGAGGAGGAGGAGTACCGCGGCGGAGTTCAGGCTGTAG 3699  
16787 CGTATGTCATTTCCACCGTGGGAGGAGGAGTATGAGGAGAGAGAGAGAGAGAGAGAG 16728  
3700 CTGCGTGAAGATTCCTGACCGAGGCGGCGCACGAGCGGCGGCGGCTTCCGCGGCGAG 3759  
16727 TTGCTGGGCGATTTCTCACCCTATGAGAAACGAAAGAGAGAGAGAGAGAGAGAGAG 16668  
3760 CAGCACTTCCGCAATGATCTTGCAGCTTCTTGAAGACCAACTACCGAGTACATGCTG 3819  
16667 CAGAGCTTACAAAGATGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16608  
3820 CCTACAGTCTTCTGAGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3879  
16607 CCGAGTATGCTGATTTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16548  
3880 AAGGCGCTGCGGAGCGGAGAGATCTGCTGCGGCGGAGT-----TCGCGGAGACG 3933  
16547 CAATAATTAAGCTGAG 16488  
3934 GCGCAG 3993  
16487 CAGGTGAG 16428  
3994 GAGGTGCTGCGGCTGCGAG 4053  
16427 AGCAGCTGCTGCGGAG 16368  
4054 CCGATGAG 4113  
16367 CGTATTCAG 16308  
4114 CAGTACCGAATCTGCGCTGCTGCGG 4140  
16307 GGTCTGCAACATGCGGAGATGTCGG 16281

RESULT 5  
US-10-085-959-54  
; Sequence 54, Application US/10085959  
; Publication No. US20030165870A1  
; GENERAL INFORMATION:  
; APPLICANT: Blattner, Frederick R.  
; APPLICANT: Welch, Rodney A.  
; APPLICANT: Burland, Valerie D.  
; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073  
; FILE REFERENCE: 960296.97648

CURRENT APPLICATION NUMBER: US/10/085,955  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: 60/242,412  
PRIOR FILING DATE: 2000-10-19  
NUMBER OF SEQ ID NOS: 255  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 54  
LENGTH: 32159  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (10755)..(10795)  
OTHER INFORMATION: unsure  
US-10-085-955-54

Query Match  
Best Local Similarity 51.2%; Score 390.4; DB 12; Length 32159;  
Matches 1086; Conservative 0; Mismatches 997; Indels 38; Gaps 6;

Qy 761 CGATGATTACTGGAAGCGCGCGATCGCGAGCTCCACCTCCCGGACGCTTCGATGA 820  
Db 10021 CGCGGCTTACTGCTGGCGGAAGCATCGACGCTTCCCGCGCGCTTCTTGGCTGG 10080  
Qy 821 AGCGCGATCATCTACCTGGAAGAGATCGGCTTCGGGCAACGAGCAATGGCTGCGT 880  
Db 10081 CCTGGAAACCGCGACGCTATGTAAGTCGTAATACCGAGCGCGCATGATTGTTCCGG 10140  
Qy 881 CGGACTCCTCGGGGTCGATTGAAGCGCGCTGTGCGGGAGCGCGGCTGACCCGACGGGCG 940  
Db 10141 CAACAACGCTGGACGCGCTTTAGAACCGGGCGGCGAGTATGGCGTGAAGCGCGACATGG 10200  
Qy 941 TCATCTGCTGCTGATTTTCCGAGGTATCGGCGCGCTGAGCGCGAGCCCGCTTACCG 1000  
Db 10201 CGCTGGCGACCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10260  
Qy 1001 TCACATTAACGCTCTTCAACCGGCTCCCGTCATCGCGCGGTAACGATACCGGGG 1060  
Db 10261 TTAACTACCTTATTCGACCGCGACGCGCTGACCCGCGCGTGGCGCGATGCTTGGCG 10320  
Qy 1061 ACTTCACGTGATGCTCTCTGACATGACACCACTGCGGACGACAGAGCTTCGAAACAC 1120  
Db 10321 ACTTACCAATATTTCTTCTGCTGGAACACGCGCTGCGATGCGATACGTCAGCAACTGG 10380  
Qy 1121 GCGCTAAGCTTTCAGAGCAAGCTGTGGAGACGATGATCACTGCGACGTAAGCGGTA 1180  
Db 10381 CGCGTA---AAACGCTCAAGTCTTACGAGAGCTGGAGCATCCGCACTGGTCGGCG 10437  
Qy 1181 TCGAGGTCAGAGAGCGCGCGGCTCTGGGATCCACGAGGCGCATTTGTTCCCG 1240  
Db 10438 TCGAATTACTCGTGAATCAAGCGCAGAGCGCTACCCCGACGGGCG---CCCG 10491  
Qy 1241 TGGTCTCAGAGCGGCTTAAACCAAGTCGTTGCTGATCACTTCTGACAGAGCTCG 1300  
Db 10492 TGGTATTTACAGCAATCTGGGCGTTCCCTTACAGCAGCGCGCAGAAATGCGCTTGG 10551  
Qy 1301 GAACCTCGGTGACACGACGACGACATCTCTGAGTCTGCTGATCATCACTTACG 1360  
Db 10552 GCGAGCGGAAATGGGCGATCTGCAAAACCGCGAGGCTGGAATGATCACTTGGCGCTTG 10611  
Qy 1361 AGCAGATGGGGGCTGCTGCTGCGGTGGGACATGCTGACGAGAGTTCGCGCGCGAC 1420  
Db 10612 AGCATACGCGAGGCTTGGCTGATGGAAGAGCAACAGACGCGCTTCTCCCGCGT 10671  
Qy 1421 TTCTGACGACATGCTGGAAGCTACGTCGTTTCTCCGCGGCTCACTGAGAACAT 1480  
Db 10672 TAGTGAAACATTTGTCAGAGCTTACGCAAGTTGATTAACAACCTCTGGATGAGAA 10731  
Qy 1481 GGGGTGAACAGGTGC-----GCTGTTGCTTCCGCGCTCCAGCTGAGAACGCGGGC 1532  
Db 10732 GCGCGCTGGGAAAGCGCTTTCAGATATTATGCCCCCAGCCGCGGATACGGAAC 10791  
Qy 1533 GAGCGAAACGAGCAACGCGCTGCTGAGCGAGCATACGCTGACAGGCGCTGTTCCGGC 1592

Db 10792 GGAATCAACACCGCGGTCGCCCATTCGGAAGGCTTGTGATGAAGCATTTCCG 10851  
Qy 1593 GCGGTGACAGACTGCCCATGACAGCTGCGGTGTGTCGCGCGCAAGAGCTTACGTA 1652  
Db 10852 TATCGCTGCAACGCGCGACGCGCTGGCGGTAAACGACATGCTTATCAGTGAATTA 10911  
Qy 1653 CGAAGGCTTTCGCGCGCTGGCGGCGGCTTGGCGCGGCGGCTGCGGACGAGGGGCGACG 1712  
Db 10912 TCATAGCTGACAGACTATGCCCCGCTTGGCGGCGGCGGCTTATCGAGTGGCGGTTCA 10971  
Qy 1713 CCCGAACAATTGTGCGGTGTGATGAGAAAGCTGGAGACAGGTTGTGCGGTTCT 1772  
Db 10972 GCCCGGCAATATAGGTGATACAGATGTCGAAAGCGCAGAGCAACTTGTGTCGTTCT 11031  
Qy 1773 GCGGTGCTCGAGTACGCGCGGCTTACGTGCGATGATGCGGACCTAACCGGCGGACG 1832  
Db 11032 GCGCGCTGCTGCGCGGCGGCGGCTTACGTTTCGCTTTCGCTGATCAGCTGCGGACG 11091  
Qy 1833 TATCAGTACCTCCCTGATCAATGATGATGAAGCTGCTGATGACGACCATGCTGGA 1892  
Db 11092 GCGGAGAAATTAAGCTGACAGCGCTCGGCTGATGCTCATTTGTGACAGACG 11151  
Qy 1893 TGGCAAACTGTTCATGCGCGCGGAGTCCAGCGGCTGCTGTAAGCGAGCGCGGCTGA 1952  
Db 11152 CAGCGCGGCTCA-----GAGCATTTCCCGCTTGGCTGGACAGGCGCATTT----- 11200  
Qy 1953 AGCGACGCGACACAGCTTCGATGATGCCATTTAGACACTTGGATCTGCGTATGT 2012  
Db 11201 ---GAGCGGAGCGGATGCGCAACCGGCTGATGACCGCGCGCGGACGCAACCGGCTTACAT 11256  
Qy 2013 CATCTAACCTCGGGATCCACAGGCTGCGCAAGGGGAGTATGATCATCGGAGTGC 2072  
Db 11257 TATCTAACCTCGGCTTACCGGTACGCGCAAGGGGATGATCTTACCGGGAGC 11316  
Qy 2073 CGTCAACCAATCTGACATCAACAGCGCTTCGAAATAGGCGCGGACAGAGGATGCT 2132  
Db 11317 GCTTAACACGCTGTCGATATCAATACCGGATACAGGTGGCGCGCATGACAGGCTGCT 11376  
Qy 2133 GCGCTCTCTGCTGAGGCTGATCTCGGTCTATGATGTGTTGGGATCTTGGCGCG 2192  
Db 11377 GCGCTCTCGCGCTTACATTTGATTTTACGACATTTTGGCGTACGCGCG 11436  
Qy 2193 GGGCGGTGATCGGTGTGCGGAGCGGCTCCAGGTGCGGATCCGCGCATTTGGGACGA 2252  
Db 11437 GGGCGCGGCTGTGATGATGATGAAATATACGCGCATCTTACAGCATGTGTGA 11456  
Qy 2253 GTTATCGAACGAGAGAGTGAACGCTGTGAACTCGGTGCGCGCTGATGCGGATGCT 2312  
Db 11497 GCTGATCAAGCGCATCAAGTCAAGCTCTGGAACAGCGTCCGCGGCTGTTGATATGCT 11556  
Qy 2313 CGTCAAGCATTTGAGGCTGCGCGGATTCGCTGATGCTTCTGCGGCTTTCCTGCT 2372  
Db 11557 GCTGACCTGTGTGAAGTTTTCGCGACGCGCAACCGCGGAAACCTGCGCGCATATGCT 11616  
Qy 2373 GAGCGGCACTGGATCCCGGTGGGCTGCTGCGAGCTCCAGGCAATCAAGCGCGGCT 2432  
Db 11617 TTCGCGACGATGATGCGGCTTGAACCTCCCGCGCTTATGCGGCTTTCGCGCACAGG 11676  
Qy 2433 GTTGTGATCACTGCGGCGGCGCAACGAACTGATCTGATGCTTTCATGCGGTTACCCGT 2492  
Db 11677 ACAATTTATCGGATGAGGCGGCGCACAGAGCGTATCTGCTGTAACGCTTGGAAAT 11736  
Qy 2493 GAGGAACGTGACCTATCTGTGGCGAGATCCCTTACGCGCGCTCGCTGCGCAACAGC 2552  
Db 11737 TCACGACGCTCCCGCGCACTGCGCTCATCTTACGCTTTCGCTTAAACCAACAG 11796  
Qy 2553 GTTCAAGCTGCTGATGAGGCGCTGAAACCGCGCGGCTGAGTTCGCGGCGCACTCTA 2612  
Db 11797 CTACCGGCTGTGATGAAACAGGCGCGGACCTGCTTACCTGAGTGGGCTGCGGATTAATG 11856  
Qy 2613 CATTTGCGGCTGCGGCTGCGACTGCGCTTACGCGGCTGATGAAGAAACGCGCAAGG 2672

Accession	Sequence	Position
Db	11857 GATTGGCCGGCATTGGGGGTCGCGGAAAGCTATTTCACAGATCCCTGGCTAGCGAGCAGCA	11916
Qy	2673 CTTCCTCGTGAACCCCGAGACCCGGGAGACCGCTCTTACAAAGCCGGCGATCTTGGGCGCGTA	2732
Db	11917 ATTTTGTACGCTCCCGGA-----CGACGCGTGGTATCGCACCCGCGCATCTTCGCTGCTTA	11970
Qy	2733 CCTCCCGCATGAAACATCGAGTTTCATGCGGGCGTGAAGACACCAATCAAGCTTCGCGG	2792
Db	11971 CTGGCCAGATGGCACAATCGAGTTCTCGGTCTGTGCGACAAAGCATAGTCTCAAGTTCGAGG	12030
Qy	2793 ATTACCGCGTTAGCTCGGGGAAATTCGAGAAACGCTCAAGTCGATTCGGAAGTACGGGA	2852
Db	12031 ATATTCGATCGAGCTGGGCGGAAATCGAAAGCGCGCTACGCCAGCTGGCGGGGGTGAACA	12090
Qy	2853 CGCGGTGATTTGGCCCGCTCGG 2873	
Db	12091 AGCAACCGTTCTGGCGGATCGG 12111	

```

RESULT 6
US-10-314-657-1
: Sequence 1, Application US/10314657
: Publication No. US20030175888A1
: GENERAL INFORMATION:
: APPLICANT: SHEN, Ben
: APPLICANT: CHENG, Yi-Qiang
: APPLICANT: TANG, Gong-Li
: TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
: TITLE OF INVENTION: Synthases and Methods of Use
: FILE REFERENCE: 054030-0021
: CURRENT APPLICATION NUMBER: US/10/314,657
: CURRENT FILING DATE: 2002-12-09
: PRIOR APPLICATION NUMBER: PCT/US02/08937
: PRIOR FILING DATE: 2002-03-22
: PRIOR APPLICATION NUMBER: US 60/278,935
: PRIOR FILING DATE: 2001-03-26
: NUMBER OF SEQ ID NOS: 214
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 1
: LENGTH: 135638
: TYPE: DNA
: ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

```

Query Match	8.5%;	Score 360.6;	DB 12;	Length 135538;
Best Local Similarity	55.4%;	Pred. No. 7.6e-84;		
Matches 768;	Conservative	0;	Mismatches 604;	Indels 15; Gaps 3;
Qy 1541	ACCGGACCAACCGCGCTGCTAGGAGCATATGCTGCAACGGCCCTGTTCCGGCGCGCGGGTGTG	1600		
Db 72583	ACACACACCGCGCGCTTCCCGACACACCGGCTGATGCAACAGCTCTTCGAGAGACCAAGCCG	72642		
Qy 1601	AGCAGCTGCCATCAGCTCAGCTCGCGGTGTGTG---GGCGCGCAAGATCGCTACGTACGAAG	1657		
Db 72843	CCACGACAGCTCCGGCGGAGGCGCTGTGCTGGCGCGCGCGCGCACAGATACATTAACAAG	727020		
Qy 1658	AGCTTTGCGCGCGCTTCGCGCGGACCTTGCGCGCGCGCGCTCGCGAGCAGGGGGCAACGCCGCA	1717		
Db 72703	AGCTGAACCGGCGGGCCCAACCGCATCCGCCGCCGCGCTGCGCGCGCGAGACGTGGGGCCCG	727623		
Qy 1718	ACAACATTGTCGCGCGTGTGATGTGAGAAAGGCTGGGAGCAGTGTTCGCGGTTCTCGCGG	1777		
Db 72763	AGACGGTCGTGCTCGCTCAGTGTGTGCGCGCGGGCCCATGATGTGTGCGGTGTCTTCGGGA	728222		
Qy 1778	TGCTCGAGTACGAGCGCGGCTTACGTGCGCATTCGATGCCCACTTACCGCGCGAGCTCATCC	1837		
Db 72823	TTCGGAAGGCGCGGCGGTGTCTATCTGCGCATTGAAGCCGATCTGCCCGCGAGCGGGCCG	728823		
Qy 1838	ACTACCTTCCTGCATCATGTGTGAGGTAAAGCTCGTGCTGACGACGACATGTGATGGCA	1897		
Db 72883	CCGTCATCTCTGAGAGAGCGCAACGCGAGGTCTGTCTCAACACCGCGCAACCGGAGGGCT	729424		
Qy 1898	AAC---TGTATAGCGCGCGGGGATCCAGCGGCTGCTGTGACGAGGCGCGGCTGGAAG	1954		

Db	72943	GGCCGGTGGCCGAGCGGATACGCCCCGGTGTGGCCGGAGCCGCCCGTGGAGGGCCCCCAACC	73002
Qy	1955	GCGACGGCGACCAAGCCTCCGATGATGCCAATTCAGACACTTCGGATCTCGGTAATGTCA	2014
Db	73003	CCGCCGACGCGGAGCAACTGGCCCCCGGCGCCGTGACGACGCCGACAACACGGGGTACATCA	73062
Qy	2015	TCTACACCTTGGGATCCACAGAGGTTGCCAAGGGGGTATGATTCATCTGGGGGTCCG	2074
Db	73063	TCTTCACCTTGGGGACAGACCGGGCGGCCCAAGGGTGTGGCGGTGGCGCACCGGCCGTTC	73122
Qy	2075	TCACACCATCTCTGGGACATCAACGAGCGCTTCGAAATAGAGGCCCGGAGACAGAGGTGTCTGG	2134
Db	73123	TGAACCTCATCAACTGTGTGCCGGCGACCTTCGGGGTTGGGGCCCCGGGAGACATGGGCTGT	73182
Qy	2135	CGCTCTCTGCGTGAAGCTTCGATCTCTGGGTCTATGATGTGTTCGGGATCTTGGCGGCGG	2194
Db	73183	GTGTCACTCTGCGCTCGGCTTTCGACTGTCTGTTGTGACAGTGTTCGGCGCTGCTCGGTAACGG	73242
Qy	2195	GCGGTATGATCTGTGTGTGCCGGAGACGGCTCCAAAGTGTGGGCAATCCGGCGCATTTGGGCAAGT	2254
Db	73243	GTGTCCGGGCTCTACATCTGGCGAGACGGGACAGACAGAGCGATCCCGTCTCTCTCTGGAGG	73302
Qy	2255	TGATCGAATCAGAGAGAAAGTGAACGGTGTGGAACCTCGGTGCGGCGCGCTGAATGCGGATGTCTGG	2314
Db	73303	TGCTGATTCAGAGAGACCGGTCATCTTCTGGAACTCCGGCGCCACACAGCTCGGCCAGATAG	73362
Qy	2315	TCGAGCATTTTGAAGGTCGCCCCGATTTGCTCGCTAGAGTCTGTGGGCTTTGCGTGTGA	2374
Db	73363	GTCCGCTGTCTGACACAGATGGGACACGGCGGGATACCGGTATCTGGCGCTGTGTTCTTCA	73422
Qy	2375	GCGGCGACGTGAATCCCGGTGGGGGCTGCTGGGAGAGCTCCAGGCCATCAAGCCCCGGGTGT	2434
Db	73423	GCGGCGACTTCACGCGCGCTGCTCCCTCCCGAGAGGTGTGGCGCGCTTTCGCGCGCGCCG	73482
Qy	2435	CGGTGATCAACCTTGGGCGGGGCCACCGGAAGCGTGCATCTGTCCATTCGGGTTACCCCGTGA	2494
Db	73483	ACATGATCAACCTTGGGCGGGCGGTCACCGAGCGACGGTGTGTGTCCACTGTGGTTCCGATTCG	73542
Qy	2495	GGAACGTGCACTTATGTGTGGGCGAGCATCCCTTACCGGCGGTGCGCTGCGCAACAGACGT	2554
Db	73543	GGGGGATCGAACCAGCCTGGCGGACGATCCCTTACCGGACAGGCCATTCGACACACCGCT	73602
Qy	2555	TTCACCGTCTCATGATGAGGGGCTTCGAACCGCGCCGGTCTGTGGTTCCGGGGGAACTTCTACA	2614
Db	73603	ACCACTGTGTGACGAGGGCGCTGGCCCCCTGTCTCGGTCGTGAGGGGATCTGTCTACA	73662
Qy	2615	TTGGCGGGGGTCTGGGCTTGGCACTTGGGCTTACGTGGGCGCATGAAAGAAAGACGCCAAGAGCT	2674
Db	73663	TGGCGGGTGTGTCTTGTGGCCTTGGCTTACGTCAACCAAGCCGAGCTGACCGCGGACCGGT	73722
Qy	2675	TCTCTC-----GTGCAACCCCGAGACCGGGGAGCGCTTCAACAAGACCGGCGATCTTGG	2725
Db	73723	TCATCCCCGACCCGTTCTCAAGAGAGACCTTCAAGGAGCGCTTCAACAAGACCGGTGACAGGG	73782
Qy	2726	GCGGCTTACCTGCCCCGATGAAACATGAGATTCATGTGGGGGTGAGAGCAACCAATCAAC	2785
Db	73783	CCCTGTATCTACCCGAGACGGCAACTAGGCTTCCAGGGGGGGGCGGACGGACGTGTGAAGG	73842
Qy	2786	TTGCGCGGATACCGCGTGTGAGCTCGGGGGAATCGAGGAACGCTCAAGTCTCGATCGGAACG	2845
Db	73843	TTCGCGGGGTTCCGGGTGTGAACGTGGCCGAATCGAGACCGGCTTGGGGGGGACAGACGGGG	73902
Qy	2846	TACGCGACCGGCGTGTGTGTGCCCCGTGGGAAACGACGGCGGAACAAGCTCTTCTAGCCT	2905
Db	73903	TCAAGGACCGCGTGTGTCTGGCCCGGGAAGACGGGTGTGGGGGACCGCAACCTCTGTGCGGT	73962
Qy	2906	ATGTGTT 2912	
Db	73963	ACCTCTGT 73969	

## RESULT 7

US-09-815-242-7601  
Sequence 7601, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trivick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: fastseq for windows Version 4.0  
SEQ ID NO 7601  
LENGTH: 1416  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(1416)  
US-09-815-242-7601  
Query Match 7.2%; Score 305.6; DB 9; Length 1416;  
Best Local Similarity 54.4%; Pred. No. 7.6e-70;  
Matches 662; Conservative 0; Mismatches 559; Indels 21; Gaps 3;  
QY 1602 GCAGCTGCCATGCAAGCTGCGCGTGTCTGCGCGCGCAAGCGCTTACGAGAGCT 1661  
DB 57 GCAGAGCGCGGCGCGCTGCGCGTATCGAGATGCGTATTCAGTGAATTATCATGAGCT 116  
QY 1662 TTGCGCGCTTGGCGCGCACTTGGCGCGCTGCGCGCGAGCGCGCGCGCGCGCGCG 1721  
DB 117 GACAGACTATGCGCGCGCTTGGCGCGCGAGTTATTCAGTGTGCGCGCTTACCGCGCGA 176  
QY 1722 ATTGGCGCGGTGTGATGAGAGAGGCTGGGAGCGAGTGTGCTGCGCGGTCTCTGCGGTGCT 1781  
DB 177 TATGTGGCTATTCAGATTCGAAAGCGCGAGCAAACTTTGTCGCTTCTGCGGTCTCT 236  
QY 1782 CGAGTCAGGCGCGCGCTTACGTGCGCGATCGAGCGCGAGCGCGCGAGCTTATCCACTA 1841  
DB 237 GCTGGCGCGGCGGTTTACGTTCCGCTTTCGCTGATTCAGCGTCCGCGCGAGGAGAA 296  
QY 1842 CCTCTCGATCATGCTGAGGTAAGCTGTCTGACGCGAGCGAGCTGAGTGGCAACT 1901  
DB 297 AATTCATGCTGACGCGCGCGCTGCGGTGCTGCTCA-----TTTGTCAACAGACCGCGAG 350  
QY 1902 GTCATGCGCGCGCGGATCCAGCGGCTGCTGTCGAGCGAGCGCGCGGCTGAGAGGCGAGG 1961  
DB 351 CGCGCGGTGACAGATATTCCTGCTTCTGCGCGAGCGAGCGCGCGCGCGCGCGAGT 410  
QY 1962 CGACCAAGCTCCGATGATCCCATTCAGACACTTTCGAGTCTGCTGATGATGATCTACAC 2021  
DB 411 CGCGCAAGC-----GTTAGTACGCGCGCGCGCGAGCGAGCGCGCTTATATATTCAGAC 461

QY 2022 CTCGGATCCACAGGTTGCCCAAGGGGTGATGATCATGATCGGGGTGCCGTCAAC 2081  
DB 462 CTCGGCTCTTACCGGTACCGGAAAGGGGTAGTCACTTTCTACCGGGAGCGCTTAAC 521  
QY 2082 CATCTGACATCAACAGAGCGCTTGAATAGGGCGCGAGACAGAGGTGTCGCGCTCTC 2141  
DB 522 CTGTTGCAATATCAATACCGGTATTCAGGTGTGGCCGATGACAGGTGTGGCCCTCTC 581  
QY 2142 CTCGCTGAGCTTCGATCTCTGCTATGATGTGTTCGGGATCTGCGCGCGCGCGGTAC 2201  
DB 582 CGCCCTTACATTTTGAATTTATCGTTTACGACATTTTTCGCTGCTGCGCGCGCGGAC 641  
QY 2202 GATCGTGTGCGCGAGCGGTCCAGTCCAGTCCGCGATTCGGCGGATTTGGGAGATTATGA 2261  
DB 642 GCTGTGATGTGATGAGAAATCAACGCGCGGATCTCACGCAATGTGTGAGCTATCA 701  
QY 2262 ACGAGAGAGGTGACGCTGTGGAACCTCGTCCGCGCGCTGATGCGGATGCTGTGAGCA 2321  
DB 702 GCGCATACGCTCAAGCTCTGGAACAGGCTCCGCGCGCTGTTCGATATGCTGTACCTG 761  
QY 2322 TTTTGGAGGTGCGCGCGATTCGCTGCTGAGCTCTCTGCGGCTTTCGCTGTGAGCGGGA 2381  
DB 762 GTGTGAAGTTTCGCGACGCGCAAGCGGAAACCTGCGCGCAATGATGCTTTCGCGGGA 821  
QY 2382 CTGATCTCCGCTGCGCGCTGCGGAGCTCCAGGCGCATCAAGCGCGCGCGGTGTGAT 2441  
DB 822 CTGATCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 881  
QY 2442 CAGCTGAGGCG 2501  
DB 882 CGCGATGCG 941  
QY 2502 CGACCTATCTGAGCGAGCATCTCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2561  
DB 942 CCG 1001  
QY 2562 GCTGATGAGCG 2621  
DB 1002 GGTGATGAACAGCG 1061  
QY 2622 GGTGCGGCTGCGCACTGCGCTACTGCGCGCATGAGAGAGAGCGCGCAAGCGCTTCTCT 2681  
DB 1062 CATTTGGGCTCGCGAGAGGCTATTTCAAGATCCCTGCGTATGCGAGCAATTTTAC 1121  
QY 2682 GCACTCCAGAGCG 2741  
DB 1122 GCTCCCGGA-----CGAGCGCTGTATTCGACACCGCGCATCTCGGCTCTACTGCGCAGA 1175  
QY 2742 TGGAAATTCAGTTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2801  
DB 1176 TGGCAATTCAGTTCCTGCGTGTGCGCGCAAGAGGTCAAGTTCGAGATATTCGAT 1235  
QY 2802 TGAGCTCGCGGAAATCGAGAAACGCTCAAGTGCATCGAACGTACCGGCGGTAT 2861  
DB 1236 CGAGCTGCGCGAAATCGAAAGCGGCTCAGCGAGTGGCGGGGTGAACAAGCAACGCT 1295  
QY 2862 TGTGCGCGTGG 2873  
DB 1296 TCTGCGGATCGG 1307  
RESULT 8  
US-10-156-761-3143  
Sequence 3143, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 PRIOR FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 3143  
 LENGTH: 1875  
 TYPE: DNA  
 ORGANISM: Streptomyces avermitilis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(1875)  
 US-10-156-761-3143

Query Match 7.0%; Score 296.8; DB 14; Length 1875;  
 Best Local Similarity 53.0%; Pred. No. 1.6e-67;  
 Matches 743; Conservative 0; Mismatches 632; Indels 27; Gaps 4;

1593 GCGGCTGAGAGAGCTGCCATGACAGCTCCGCTGTGTCGCGCGGCAAGAGCTCAGCTA 1652  
 132 GCAGGTGCTCCGAGACCTTGAGCCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 191  
 1553 GAGGCGAAACGCGAACGCGCTGCTGAGCGAGCATGCTGCAAGGCTGTTGCGGCGC 1592  
 72 GAACGACACCGCGGAGCTGCGCGGCTGCGGACACCTGCGGAGCTGTTCCGTAAC 131  
 1593 GCGGCTGAGAGAGCTGCCATGACAGCTCCGCTGTGTCGCGCGGCAAGAGCTCAGCTA 1652  
 132 GCAGGTGCTCCGAGACCTTGAGCCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 191  
 1553 GAGGCGAAACGCGAACGCGCTGCTGAGCGAGCATGCTGCAAGGCTGTTGCGGCGC 1712  
 192 CACCGAGCTGACGAGCGCTGAGACCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 251  
 1713 CCGGACACATTTGCTGCGGTGCTGAGTGAAGAAAGCTGGAGAGAGTTGTCGCGGTTCT 1772  
 252 CCGGACCGGCTGCTGCGCTGACCGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311  
 1773 CCGGCTGCTGAGTCAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1832  
 312 CCGGCTGCTGAGTCAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371  
 1833 TATCCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1892  
 372 CATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 431  
 1893 TGGCAACTGT-----CATGCGCGCGCGGAGTCCAGCGGCTGCTGCTGAGCGAGC 1943  
 432 GCGACGCTGCG 491  
 1944 CCGGCTGAGAGCGACG 2003  
 492 GGAAGACCGCTGAGCG 551  
 2004 CCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2063  
 552 CCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611  
 2064 TCGGCGTCCGCTCAACACCATCTGAGCATCAAGAGCGCTTGGAAATAGGCGCGGAGGA 2123  
 612 CCGCGCATAGTCAACCGCGCTGCACTGATGACAGACAGATACCGGCTGACCGCGAGAGA 671  
 2124 CAGGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2183  
 672 CCGGCTGCTGAGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 731  
 2184 CCGTGGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2243  
 732 GCTGATACCGCTA 791  
 2244 TTGGGCAAGAGTTGATCGAACGAGAGAGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2303

Db 792 CTGGCCGAGCTGATCCGCGGCGAGCGAGATCAACACCGTCACTTCTCCGCTGATGCT 851  
 Oy 2304 GCGGATGCTGCTGAGCATTTTGAAGGCTGCGCCCGATTTGCTGCTGCTGCTGCTGCTGCT 2363  
 Db 852 CAGGCTTCT-----CTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 902  
 Oy 2364 TTGCTGCTGAGCGGCGAGCTGATTCGCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2423  
 Db 903 GGTGTGTGAGCGGCTGAGCGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962  
 Oy 2424 GCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2483  
 Db 963 GCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019  
 Oy 2484 GTACCCCGTGAAGAGTGAAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2543  
 Db 1020 GTACTGGAAGTACACCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1079  
 Oy 2544 CAACGAGACGTTCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2603  
 Db 1080 GAACACCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139  
 Oy 2604 GCACTTACATTTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2663  
 Db 1140 CAGCTGTATCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199  
 Oy 2664 GCGCAAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259  
 Db 1200 GCGGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259  
 Oy 2718 GATCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2777  
 Db 1260 GAGACTGCG 1319  
 Oy 2778 AATCACTTTCGCGGATACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2837  
 Db 1320 GGTCAAGATCCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1379  
 Oy 2838 TCGGAAGTACGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2897  
 Db 1380 TCGCGGCTTTCGCGGCT 1439  
 Oy 2898 TCTAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2919  
 Db 1440 GGTGGCTATGTGTCTCCGCGG 1461

RESULT 9  
 US-10-156-761-1/c  
 Sequence 1, Application US/10156761  
 Publication No. US20030119018A1  
 GENERAL INFORMATION:  
 APPLICANT: OMURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 PRIOR FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 1  
 LENGTH: 9025608  
 TYPE: DNA  
 ORGANISM: Streptomyces avermitilis  
 FEATURE:

NAME/KEY: misc feature  
LOCATION: (4187715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match  
Best Local Similarity 53.0%; Pred. No. 7.8e-67;  
Matches 743; Conservative 0; Mismatches 632; Indels 27; Gaps 4;

```
QY 1533 GAGCGCAACGCGACCAACGCGCTGTGAGCAGATACGCTGACAGCGCTTTCGCGC 1592
DB 3932113 GAACGACACCGCGCGGAGCTCGCGCGGCGAGCACCCTGCGGACTGTCGTCAC 3932054
QY 1593 GGGGGTCGAGCAGCTGCCATGACAGTGGCGGTGTGTGGGGCGGAAACCTCAAGTA 1652
DB 3932053 GAGGGTTCGCGGACCCCTGAGCGCAACCGCTGTGTTCGAGAACCCGCTCAAGCTA 3931994
QY 1653 CGAAGAGCTTTCGCGCGCTTTCGCGCGACTTGGCGCGCGCTGCGGAGCAGGCGGACAG 1712
DB 3931993 CACCGAGCTGAGCAGAGCGCTCGGACCGCTTGGCGCGCGCTGCGGAGAACTGCTGCT 3931934
QY 1713 CCGGACACATTGTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1772
DB 3931933 CCGGACACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931874
QY 1773 CGCGGTCTGAGTCAAGCGCGCGCGCTTACGTCGATGCGGACCTACCGCGCGAGCG 1832
DB 3931873 CGCGGTCTGAGTCAAGCGCGCGCGCTTACGTCGATGCGGACCTACCGCGCGAGCG 3931814
QY 1833 TATCCACTACTCTCTGATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1892
DB 3931813 CATCGGTACTCTCTCTGCGGACCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGT 3931754
QY 1893 TGGCAAACTGT-----CATGCGCGCGCGGAGTTCAGCGCGCTGTGTGTGTGTGTGT 1943
DB 3931753 GCGGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGTGTGTGTGTGT 3931694
QY 1944 CGGCGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931634
DB 3931693 GGAAGAGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931634
QY 2004 CGCGTATGTATCTATCACTCTCGGAGTTCACAGGCTTTCGCGGAGGAGTGTATGATCA 2063
DB 3931633 CGCGTATGTATCTATCACTCTCGGAGTTCACAGGCTTTCGCGGAGGAGTGTATGATCA 3931574
QY 2064 TCGGGGTGCGGTCAACACATCTGTGACATCAAGCGCGCTTGTGAATAGGCGCGGAGA 2123
DB 3931573 CGCGGCTATGATCAACCGCGCTGACATGATGACAGTACCGGCTGACCGCGGACGA 3931514
QY 2124 CAGGGGTGCGGGCTCTCTCTGCTGAGCTTTCGATCTGCGGTGTGTGTGTGTGTGTGTGT 2183
DB 3931513 CCGCGTCTGCAAGAGACCGCTCTCGGCTTTCGAGCTGTGTGTGTGTGTGTGTGTGTGT 3931454
QY 2184 CCGGCGCGCGCGCGGTACGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2243
DB 3931453 GGTGATCAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTA 3931394
QY 2244 TTGGGCAAGTTGATGAAAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2303
DB 3931393 CCGTGGCGGAGCTGATCCGCGGCGGAGCGGATCAACCGCTGCACTTGTGTGTGTGTGT 3931334
QY 2304 GGGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2363
DB 3931333 CCGAGTGTTC-----CTGAGCAGCGCGCGCGCGCGCGCTTGTGACCGCGCGCTGCGCG 3931283
QY 2364 TTTCGCTGTGAGCGCGGAGTGTATCCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2423
DB 3931282 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931223
QY 2424 GCGCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2483
DB 3931222 GCGCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931166
```

```
QY 2484 GTAACCGGTGAGGAACGTGTGACCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2543
DB 3931165 GTACTGGAGATGACACCCCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTG 3931106
QY 2544 CAACGAGCGTGTTCACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2603
DB 3931105 GAACACCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931046
QY 2604 GCACTTCAATTTGGCGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2663
DB 3931045 CGAGCTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930986
QY 2664 GCGCAAGAGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2717
DB 3930985 GCGGAGACGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930926
QY 2718 CGATGTGGCGCGCTACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2777
DB 3930925 GGAACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3930866
QY 2778 AATCAAGCTTTCGCGGATACCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2837
DB 3930865 GGTCAAGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3930806
QY 2838 TCCGAACGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2897
DB 3930805 TCCGCGGTCTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3930746
QY 2898 TCTAGCCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2919
DB 3930745 GGTGCGCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930724
```

## RESULT 10

```
US-10-156-761-3184
; Sequence 3184, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3184
; LENGTH: 5304
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5304)
US-10-156-761-3184
```

Query Match  
Best Local Similarity 52.0%; Pred. No. 1.5e-60;  
Matches 753; Conservative 0; Mismatches 656; Indels 39; Gaps 5;

```
QY 1503 GTTTCGCGTGTGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1562
DB 741 GCGACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800
QY 1563 CGAGCATACGCTGACGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1619
```

Db 801 GGAACCGTGGCCCTCGACGAGCTTTACCCCGGAGGGCGGCTGACCCCGGAGCCAC 860  
Qy 1620 CGCCCTGTGTGTGGCGCCCAAGACCTGACGTAAGAGCTTTGCGCGCTTTCGCGCG 1679  
Db 861 CGCCGTGTGTGGCGCCCAAGACCTGACGTAAGAGCTTTGCGCGCTTTCGCGCG 920  
Qy 1680 ACTTGGCGCGCGCTTGGCGAGCAGGAGGAGCAGCCGAAACATTGTTGCGCGGTGTAT 1739  
Db 921 GCTGGCGCGAGCGCTGGCGAGCGGGGGCGGCGCAGCAGCTGTGCGCGGTGTGAC 980  
Qy 1740 GGAAGAAAGCTGGGAGAGTGTGCGCGTTTCGCGGTGTGCTGAGTACAGGCGCGCTTA 1799  
Db 981 GAGCGCTGCGCGAGCTGTGTGCGCGTACCGCTTCAAGAGCGCGGCGCGCTTA 1040  
Qy 1800 CGTGGCGATGATGCGGAGCTTACCGCGGAGGAGTTCATCTACTCTGATGATGTGA 1859  
Db 1041 TCTGCGGTGAGACCCCGGCTACCCGCTGAGGAGTACCGCTACGTCCTGAGAGACAGCGC 1100  
Qy 1860 GGTAAAGCTCGTGTGAGCAGCAGCCATGGCTGATGGCAAACTGTATGGCGCGCGGAGT 1919  
Db 1101 CGCTGTCTCGCTCAGACAGCCCGCGCACCGGAGCTT-----GATCAGAGAGG 1151  
Qy 1920 CCAAGCGGTGCTCTGAGCAGAGCGCGGCTGCAAGCGGAGCAGAGCTTCGATGAT 1979  
Db 1152 CCGGAGGTACTGACCTGACCGAGAGTCCGCTACCAAGAGCAGACAGCGCGCGA 1211  
Qy 1980 GCCCATTCAGACACCTTGCAGTCTGCGGTATGTCTACCTCGGGATCCAGAGGTT 2039  
Db 1212 GGTATTCGGCCAGCGCGCGGAGCTGCTGCTACTTACACTCGGCTGAGCGCGCA 1271  
Qy 2040 GCCCAAGGGGTGATGATGATCATGCGGGTGCCTCAACACCATCTGAGCATCAACGA 2099  
Db 1272 TCCCAAGGGCGCTCAGATGAGCAGCGCTGCGTGTCAACCGCATCACTGATCAGCG 1331  
Qy 2100 GCGCTTCAAAATAGGGCCCGGAGAGAGAGGTGTGCGCTCTCTGCTGAGTGTGATCT 2159  
Db 1332 CGCTTACCCCGCTCAGCTCGCGAGTATGAGAGAGAGCGCCACCTCTTCGAGCT 1391  
Qy 2160 CTCGGTATGATGATGTGTGCGGATCTTGGCGGCGGAGTACGATGTTGTCGCGAGCG 2219  
Db 1392 GTCCGTCTGGAGAGCTTCTGTGTGATGTTGAGAGGGCCCGCGGTGTGCTGCCAGCC 1451  
Qy 2220 GTCCAAAGTGCAGATCCGCGCATTTGGCGAGAGTGTGATGAAAGAGAAAGTTCAGCT 2279  
Db 1452 CGGCGCGAGAGCGGAGCGCGAGGCGCTGTACCGCATGAGCGGTACCGGTTGACAC 1511  
Qy 2280 GTGAAATCTGCTGCGCGCGCTGATGCGGATGCTGTCGAGCACTTTTGGGCTGCCCGCA 2339  
Db 1512 CATGCACTTGTGCGCTGATGCTGAGTGGCGCTTCTGCACTTGTGTTGGAGCGGCGCT 1571  
Qy 2340 TTCGCTGCTAGAGTCTTGGCGCTTTCGCTGCTGAGCGGCGCATGAGATCCGCGTGGCT 2399  
Db 1572 ACAGGCGCGGCTGTGTGTGCTGCGCAGAGTCTTCCGCGAGCGGTAGCGCTCACCGCA 1631  
Qy 2400 GCTTGGCGAGCTCCAG-----CATCAGGCCCGCGGTGTGCTGAT 2441  
Db 1632 CCAGGCTCGGCGCTTTCGCGAGGTGCGCGCGCACACAGGCGCGCGCGCTGAT 1691  
Qy 2442 CAGCTTGGCGGCGGCGCACCGAGAGCTGATGTGTCTCATGCGGTATCCCGTGAAGAGT 2501  
Db 1692 CAACCTGTAGCGGCCCGCACCGAGGCGACCGTGAAGTCCACACCGGCTGGCGCGCA 1751  
Qy 2502 CGACCTATGTTGGGAGAGATCCCTACGCGCGGTGCGGTGCGCAACGAGCTTCAGCT 2561  
Db 1752 CGAGCT-----GAGCGGGTTCGATGCGCGCGCGCATGACACACCGCATCTCTGT 1805  
Qy 2562 GCTGATGAGGCGCTGAGACCGCGCGCTGTGGGTTTCGCGGCAACTTCAATTATGGCG 2621  
Db 1806 CCTGCACTCCGAGTGGCGCGCGCGCGCTGTGTGCTGCGCGCAACTGTGATGCGCGG 1865  
Qy 2632 GGTGGGTTGCGACTGGGCTACTGGCGCGATGAAGAGAGAGCGCGCAAGAGTTCCTGT 2681  
Db 1866 TGTGCTGTGGCGCGGCTTACTGCGCGCGGAGTGAACCGCGGAGGCTTGTGCTG 1925

Qy 2682 GCACCCCGAGACCGG---GAGCGCTCTTACAGACCGCGATGTGGCCGCTACTGCC 2738  
Db 1926 CGCCCGCGCGCGGTGAGAGCGCGGTCTACGACCGGCGGACCTGGCGGTGCTGCC 1995  
Qy 2739 CGATGAAACATCGAGTTTATGCGGCGGTGAGAGACACCAATTAAGTTCGCGATACG 2798  
Db 1986 CGAGCGGAGCATTCATCACTTGGCGCGGTATGACACACAGGTCAAGGTGCGCGCTACCG 2045  
Qy 2799 GCTTGGAGTGGGGAAATTCAGAGAAAGCTCAAGTCCGATCGCAAGCGGAGCGCGGT 2858  
Db 2046 CATGAGCTGCGGAGATGAGAGTGTGCTGCGCGCGCGCGCGCGAGAGAGCGCT 2105  
Qy 2859 GATTGTGCGCTCGGAGAGCAGCGCGGAGACAGCTCTTCTAGCTTATGTGTCCGGA 2918  
Db 2106 GTGTGCTGCGCGGAGAGCGCGGAGAGCGGTACAGCCACTGTGGCGCTACGTCAAGTGGC 2165  
Qy 2919 GGGGACAC 2926  
Db 2166 GAGCGCC 2173

RESULT 11  
US-09-815-242-4052  
Sequence 4052, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4052  
LENGTH: 13029  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-4052

Query Match 6.2%; Score 261.4; DB 9; Length 13029;  
Best Local Similarity 51.8%; Pred. No. 4.1e-58;  
Matches 704; Conservative 0; Mismatches 631; Indels 24; Gaps 4;

Qy 1572 GCTGCAAGCGCTGTTCGCGGCGCGGTGAGCAAGCTCCCATGCAAGCTTGCCTGTGTC 1631  
Db 3387 GCTCCGAGAGCTGTGAGGCGGCAACTGGCGCACTCCGCGAAGCGGTGCTGAGTG 3446  
Qy 1632 GGGCGCAAGCGCTTACAGTACAGAGCTTTGCGCGCGCTTGGCGGAGCTTGGCGCGG 1691  
Db 3447 GGAAGGCGGAGCTTGGCTACCGGAGTGAAGTGAAGCGCGCGCGCAACCGCGCTGCACTA 3506



```

QY 1692 GCTGCGGAGCAGGGGGCAGCCCGCAACACTGTGTCGGGTGTGATGAGAAAGCTG 1751
DB 3507 CCTGGCGCGACAAGGGGGTGGGCCCGGACCTGGGGGTAGGATCTGGCCGGTTCGCG 3566
QY 1752 GGAGCAGGTTGTGCGGGTCTCGCGGGTGTGATGAGGGCGGGGCTTACGTGCCATGCA 1811
DB 3567 GCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3626
QY 1812 TCCCGACCTTACCGGCGAGCGATTCACATCACTCTCTGATCAATGTGAGTAAAGCTGT 1871
DB 3627 CCGGACCTACCCAGGAAAGCGCTGGCTTACATGCTCGCGACAGTGGGGTGTGAGTGT 3686
QY 1872 GCTGACGACGCCATGTGCTGATGCGCAACTGTGTATGCGCGGGGATCCAGCGCTGT 1931
DB 3687 ACTGACCCAGGCGCCATCTGTTCGACAGCGCTGTGCGGGGAGGGGCGTCAACGCCGATCTG 3746
QY 1932 CGTGAAGGAGGCGGGGTGTGAAGGGCAGCGGCGACCGCTCGATGATGCCATTGACAGC 1991
DB 3747 CCTGACAGCCTTACAGCTGACCACTGGCCGAGCCAGGACCGGGGCTTGACCTTGACCG 3806
QY 1992 ACCTTCGATCTCGCGTATGTCTATCACTTCGGGATCCAGGGGTTGCCAAGGGGCT 2051
DB 3807 CGACA---ATCTGCGCTAGCTATCACTTCGGGCTCAGCGGCGACCGAAGGGGCT 3863
QY 2052 GATGATCATCATCGGGGTGCGCTCAACACATCTCTGACATCAAGCGCTTGCAT 2111
DB 3864 GGGCAACACCCACGCGCGCTGCGCCAGGCGCTGCAATGACAGCCACCTACACCT 3923
QY 2112 AGGGCGCGGAGCAGGGGTGCTGCGCTCTCTGCGTGTGATGATCTCTGCGGTCAAGA 2171
DB 3924 CAGCGGCGAGAGTCTGTATGAGAAAGGCGCGGTGAGCTTGTGAGTGTGTGGA 3983
QY 2172 TGTGTGCGGATCTGCGCGCGCGGCGTACGATGATGTCGCGGACGCGTCAAGCTGCG 2231
DB 3984 GTGCTTCTGGCGCGTGTCAACGCGTGTGCGGTGTGCTCTGCGCGCCCGGCGAGCACCG 4043
QY 2232 CGATCCGCGGCGATTTGGGCGAGAGTTGACAGAGAAAGGTGACGCTGTGAACTCGT 2291
DB 4044 CGATCCGCGCGCGCTGTGTGGAATGTGTGCGCGAGTTCAGCCCTGTGACCTTCTGT 4103
QY 2292 GCGCGGCGTGTGATGCGGATGCTGTGAGCATTTTGAAGGTCGCGCGGATTCGCTGCTAG 2351
DB 4104 GCGCGCGCTGTGCACTGTTATC-----GACGAACCGCGGCTGTGCGCGCTGTGCGG 4154
QY 2352 GTCTCTCGGCTTTGCTGCTGAGCGCGGACGATGATCCGCGTGGCGCTGTGCGGAGCT 2411
DB 4155 CAGCTTGGCGCGCTGTGTTCTCCGCGCGGAGGCGCTTCCGCGGAGCTGTGCGCAACCGCGT 4214
QY 2412 CCAAGCGCATCAGGCGCGCGGTGTGATGACCTGGGGGCGGCGACCGAAGCTGTGAT 2471
DB 4215 GTTGCAACGCTGCGCGCGGTGCGCTGTGATTAACCTTACGAGCCACCGAAGACCGCCAT 4274
QY 2472 CTGCTCATCGGGTACCCCGTGAAGAGTGTGATGATGAGGAGCATCCCTTACCG 2531
DB 4275 -----CAACGTACCCATTTGGCAGTGTGCGCGGAGAGCGGAGCGTTCGCTGATCGG 4328
QY 2532 CCGTCCGCTGCGCAACAGAGCTTTCAGAGTGTGATGAGGCGCTGCAACCGCGCCGCT 2591
DB 4329 CGACCGGCTGTGGAGAGTGTGTGCGCGGTGTGAGAGCGCGGATTAACCTGTGTCGCGG 4388
QY 2592 CTGCGGTTCCGGGCAACTCTACATTGCGGGGTGTGGGCTGTGGCTGTGAGCGCA 2651
DB 4389 CGCGGTGCGCGGAGCTGTGATCGCGGCTGTGGGCTGTGGGCTGTGAGCGCAACTGTGGGGCG 4448
QY 2652 TGAAGAGAGAGCGCGAGAGCTTCGCG-----TGACACCGCGAGAGCGGGAGCGCT 2705
DB 4449 CCGCGGCTGTAGCGCGAGCGCTTCTGTCGCCATTCGCTTTCGCGCGCGCGCGAGCGCT 4508
QY 2706 CTACAAAGCGGCGATCTGGGCGCTACCTGCGGATGAGAAACATGAGTTCAATGGGGCG 2765
DB 4509 GTACCGCACCGGCGAGCGCGGCTGTGAAACGCGGAGCGGTGTGGAATACCTCGGCGG 4568

```

```

QY 2766 TGAAGCAACCAATACAGCTTTCGCGATACCGGCTTGAGCTGGGGAATCGAGGAAC 2825
DB 4569 CTTGACCAAGAGCTTACAGCTGCGGTCTGCGATTCAGAGCGGAGAAATTCAGGACG 4628
QY 2826 GCTCAAGTGTGATCCGAAGTACGCGACGCGGTATGTGCGCGGTGCGGAACGACGCGC 2885
DB 4629 CTTGTGGCGGACCGGGGGGTGGCCAGCGGTATGTGATTCGCGGAAGCGTGGCGG 4688
QY 2886 GACCAAGCTTCTTACGCTTATGTGTCCCGAGGGGAC 2924
DB 4689 CAGCGATTTGTGTGCTACTACCGCGGCTGTGCGCGC 4727

RESULT 12
US-10-156-761-3629
; Sequence 3629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3629
; LENGTH: 11058
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(11058)
US-10-156-761-3629

Query Match 5.8%; Score 246.2; DB 14; Length 11058;
Best Local Similarity 51.5%; Pred. No. 3,7e-54;
Matches 709; Conservative 0; Mismatches 638; Indels 30; Gaps 5;

QY 1574 TGCAGGCGCTGTTCGCGCGCGGTGAGCAGCTCCCATGCACTGCGCGGTGTGCG 1633
DB 1469 TGCAGAGGTGTTCAGAGAGAGTGTGCGCGGACCCCGAGGGGGTCCGGTCTCTTGC 1528
QY 1634 CGCGCAAGCGTCACTGACGAAGAGCTTTCGCGCGTTCGCGGACTGTGCGCGCGC 1693
DB 1529 AGGGAGCGATGTGACGTACGCGGAGCTGAACGCCCGCCCAACCGGATTCGCCACGCGC 1588
QY 1694 TCCCGAGCAGGGGCAAGCCCGAACACATTGTGTGCGGTGTATGAGAAAGGCTGGG 1753
DB 1589 TGAATGCAAGGGGGTGTGCGCGCGGAGAGCTGTGTGAGCTGTGAGCGCGGCAATCG 1648
QY 1754 AGCAGTTGTGCGGTTCTGCGGTGTGATGAGTCAAGCGCGGCTTACGTGCGCATGATG 1813
DB 1649 AACTGATCCCGGCTCTGTGCGCATCTTAAGTCCGGGTGCGCGATGTGCGCGTGAAC 1708
QY 1814 CCGACCTTACCGGCGAGCGATATCACTTCTGATCATGTGTAGGTAAAGCTGTGC 1873
DB 1709 CGGCAACCGGCGCAGCGGATTCGCTATGTGTGAGGAGCCAGACCGCCCGGTGTGG 1768
QY 1874 TGAAGCAGCATGTGTGATGCGCAACTGTATGAGCGCGCGGAGTTCAGAGCGGTGTG 1933
DB 1769 TGAC-----CACCGCGAGCAGCGGATTTGTTCAGCTGGA-GCTGCTGCTCTTG 1819
QY 1934 TGAGCAGGCGCGGCTGCAAGCGGCGACGAGCTTCGATGTGATTCATTGACAGAC 1993

```



```

Db 1820 ACACGAGCGCCGCGGACCTGCGCGCGCGGACCTCCGACCCCGGCGCTGCCCGGACGCC 1879
Qy 1994 CTTCGATCTGCGGATGTCATCTACCTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2053
Db 1880 CCGAAGACCTGATCTACAGATCTACCTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1939
Qy 2054 TGATGATCTGCGGATGTCATCTACCTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2113
Db 1940 CCTTCACCCACACCAACCTGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1999
Qy 2114 GCGCGGAGACAGGGTGTGCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2173
Db 2000 CCGAAGACCTGATCTACAGATCTACCTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2059
Qy 2174 TGTTGCGGATCTGCGGATGTCATCTACCTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2233
Db 2060 TGTGGGGTGTGCTGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2119
Qy 2234 ATCCGCGGATCTGCGGATGTCATCTACCTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2293
Db 2120 CCGCGGAGAGTTCCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2179
Qy 2294 GCGGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2353
Db 2180 CTTCCGCTTCCTGCTGCTGCGGATCCACAGGGTTGCCAAGGGGGTGA 2239
Qy 2354 CTTCGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2413
Db 2240 TCGCGGAGAGTTCCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2299
Qy 2414 AGGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2473
Db 2300 ACCGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2347
Qy 2474 GGTTCGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2530
Db 2348 ACACCACTTACCACTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2407
Qy 2531 GCGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2590
Db 2408 GCGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2467
Qy 2591 TCTGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2650
Db 2468 TCGGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2527
Qy 2651 ATGAAGAGACAGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2704
Db 2528 GCGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2587
Qy 2705 TCTACAGAGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2764
Db 2588 TCTACAGAGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2647
Qy 2765 GTGAGAGCAACCAATCAAGCTTCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2824
Db 2648 GATGAGAGCAACCAATCAAGCTTCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2707
Qy 2825 GCGTCAAGTCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2884
Db 2708 GCGTCAAGTCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2767
Qy 2885 CGAAGCAAGCTCTTCTGATGTCGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2941
Db 2768 GCGAGAGCGCTCTGCTGCTACAGAGCGCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2824

```

```

RESULT 13
US-09-940-316B-1
; Sequence 1, Application US/09940316B
; Publication No. US20030175901A1
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.

```

```

; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTU, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE FKBP GENE OF THE PK-520 POLYKETIDE SYNTH
; TITLE OF INVENTION: GENE CLUSTER
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940.316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275) ... (71465)
US-09-940-316B-1

```

```

Query Match 5.7%; Score 241.6; DB 12; Length 77536;
Best Local Similarity 49.7%; Pred. No. 8.5e-53;
Matches 703; Conservative 0; Mismatches 699; Indels 12; Gaps 3;

```

```

Qy 1530 GCGAGCGCAACCGCGCAACCGCGCTGTCAGGAGATACGTCGACCGGCTGTCG 1589
Db 49194 GGTGCTTCGACACACACGCGCGGACCTGCTGTCGACGCGGCTGTCG 49253
Qy 1590 GCGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1649
Db 49254 CCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49313
Qy 1650 GTACGAAGCTTTCGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1709
Db 49314 CTACGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49373
Qy 1710 ACGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1769
Db 49374 GCGACCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49433
Qy 1770 TCTGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1829
Db 49434 GGTGCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49493
Qy 1830 GCGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1889
Db 49494 GCGACCGGATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49547
Qy 1890 GATGAGCAACTGTCATGCGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 1949
Db 49548 GGTTCATCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49607
Qy 1950 CGAAGCGAGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2009
Db 49608 GGAACCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49667
Qy 2010 TGTATCTACCTGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 2069
Db 49668 GCGATCTACACGTCGCGGATGTCGCGGATCCACAGGGTTGCCAAGGGGGTGA 49727
Qy 2070 TCGCGTCAACATCTGTCGATCAACAGAGCTTGGAAATAGGCGCGGATGTCG 2129
Db 49728 GCGCGTCAACCTGCTGCTGTCGATCAACAGAGCTTGGAAATAGGCGCGGATGTCG 49787

```



```

; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (78110)..(76449)
; OTHER INFORMATION: ORF 24; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (82316)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (82587)..(8446)
; OTHER INFORMATION: ORF 29; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness, N-terminus only
US-09-976-059-1

```

```

Query Match      5.6%; Score 239; DB 10; Length 88421;
Beet Local Similarity 50.4%; Pred. No. 4.2e-52;
Matches 729; Conservative 0; Mismatches 695; Indels 23; Gaps 5;

```

```

QY 1525 GCGCGGCGGAGCGCAACGCGCAACGCGCTGTGAGGAGCATACGCTGACGGGCTG 1584
DB 68266 GCAACCGGACCGGCACTGCGCGCGCGCGCGCGCTGCGCGCTGCGCGCGCTG 68325
QY 1585 TTCCGCGGCGGCGTGCAGCAGCTGCGCATGACCTGCGCGTGTGTGCGCGCGCAAGC 1644
DB 68326 TTCCGCGGAGAGTGTGCGCGCTGTGCTCCACGCGCGTGTGCGCGATGCGGAGTGG 68385
QY 1645 CTCACGTACGAAAGACTTTTCGCGCGCTTGCAGCGCACTTGGCGCGCGCTGCGCAAGC 1704
DB 68386 CACACCTTACCGGACCTGCAAGAGTGTGCTCCGCGCGCTGCGCGCGCTGCGCAAGC 68445
QY 1705 GGGGCAACGCGCAACATGTTGTCGCGGTGTGTGATGAGAAAGCTGGGAGAGAGTTGTC 1764
DB 68446 GCGGCGCGGAGCGGAGCGCGGTGCGCGCTGTGCTGACCGCTGCGCGCGAGCTGTGCGC 68505
QY 1765 GCGGTTCTCGCGGTCTGAGTCAAGCGCGCGCTACGTCAGTCAGTACGCACTACCG 1824
DB 68506 GTGCTCTCTGCGGTCTCAAGCGCGCGCGCTGTGTGTCGCTGACGCGCGCTGACCG 68555
QY 1825 GCGGAGCGTATCACTACTCTCTGATCATGTGTGAGTAAAGCTGTGTGTGACGACCA 1884
DB 68566 GCGGACCGGATGCTTACACCTCTGCGCGAGCGCGCGCTGCGGTGTGTGTGACACCTG 68625
QY 1885 TGGCGATGAGCAACTGTGATGCGCGCGCGGAGTCAAGCGCGCTGTGTGTGAGAGGCG 1944
DB 68626 GCGGACTTTTCCGAGGTGAGCGGTGTGCGGTGTGCGCGGAGAGCTGTGCGCGAGGCG 68685
QY 1945 GCGGTCGAGAGCGAGCGCGAGCGCGCTTCCGATGATGCCATTCAGACCTTCCGATCTC 2004
DB 68686 GCGGAGAGCGCGCGGAGCTC-----CGGCTGTGCGACCGCGCGCGCGAGGCGCG 68736
QY 2005 GCGTATGTATTTACACTCTCGGAGTTCACAGGAGTTGCCAGAGGGGTGTATGATCAT 2064
DB 68737 GCGTATGTATTTACACTCTCGGAGTTCACAGGAGTTGCCAGAGGGGTGTATGATCAT 68796

```

```

QY 2065 CGGGGTGCGCTCAACACCATCTGAGCATCAACGAGCGCTTCAAAATAGGGCCGAGAC 2124
DB 68797 GCGAACGTGTGCGCGGTCTGAGCGCGAGCGCGAGAGTACGCGCTGGCCCGCGAC 68856
QY 2125 AGGTGTGCGGCTCTCTGCGCTGAGCTTGCATCTCTGCTGTATGATGTGTGCGGATC 2184
DB 68857 GTGTGAGACTTCTTCACTCGGCGCGCTTGCATCTCTGCTGTGAGATGTGCGGCTGC 68916
QY 2185 CTGCGCGCGCGCGAGCATGTGTGTGCGGAGCGCGTCCAGCTGCGGAGTCCGCGCAT 2244
DB 68917 CTGCTACCGCGCGAGCATCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 68976
QY 2245 TGGGAGAGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
DB 68977 TTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69036
QY 2305 CGGATGCTGTGCG-AGCATTTTGAAGGTGCGCGCGGATTCGCTGCTAGGTCTGCGGCT 2363
DB 69037 ACGGAGCTGTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69096
QY 2364 TTCCGTGTGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2423
DB 69097 TTCCGTGTGAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69156
QY 2424 GCGCGCGCTGTGCGGTATCAAGCTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2483
DB 69157 GCGC-GCTGCGCGCGCTGTGCAATGTTTGGCATACAGAGAGAGAGAGAGAGAGAGAG 69215
QY 2484 GTACCGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2543
DB 69216 GATGAGAGTCAAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69275
QY 2544 CAACGAGAGAGTTCACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2603
DB 69276 GCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69335
QY 2604 GCAACTTACATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2663
DB 69336 GCAATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69395
QY 2664 GCGCAAGAGCTTCTCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2723
DB 69396 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69452
QY 2724 GCGCGCTACCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2783
DB 69453 GCGCGGCTGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69512
QY 2784 GCTTGCAGATACCGCGTGTGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2843
DB 69513 GCTGCGCGGCTTCCGAGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69572
QY 2844 GGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2894
DB 69573 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69632
QY 2895 CTTTACCTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2954
DB 69633 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69692
QY 2955 CCTCAAG 2961
DB 69693 CCGAGATG 69699

```

```

RESULT 15
US-10-156-761-3630
; Sequence 3630, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO

```

```

APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 3630
LENGTH: 5127
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(5127)
US-10-156-761-3630

Query Match
Best Local Similarity 51.0%; Score 233.4; DB 14; Length 5127;
Matches 692; Conservative 0; Mismatches 641; Indels 24; Gaps 5;

QY 1559 TGAAGCAGATACGCTGCAAGCGCTTTCGCGCGCGGGTCGAGCAGTCCCAACAC 1618
DB 2 TGAAGCCGCAACGCTTTCGGAACGCTTCGCGAGAGGGGTTCCGTACGCGCACGCG 61
QY 1619 TGGCCGTGTGTGCGCGCGCGCAAGCGCTACCTGACGAAGCTTTGCGCCGTTCCGCG 1678
DB 62 CGCCGCTGTGCAAGGAGCGCGCGAGCGGTACGTACGAGAGCTGACCGTGCCTCC 121
QY 1679 GACTTGGCGCGCGCTGCGCGAGCAGAGGAGCAGCCGGAACACATTTGTCGCGGTGTA 1738
DB 122 GCATGCGCGCGCACTGCGCGGACTTGGGTGCGGGGCGCGAGAGCTGTCGGGTAGCC 181
QY 1739 TGAAGAAAGCTGGAAGAGAGTTGTGCGCGTTCTCGCGTCTGAGTCAAGCGCGCT 1798
DB 182 TGCCTGGCGCGCGTGTGATGTGTCGCGCTGCGCTGCGCGTGGCGGGCGGGGCT 241
QY 1799 AGTGGCGATTCATGCGCGACCTACCGCGAGGTATCCACTTCTCTGATCATGTG 1858
DB 242 ATGTCCGCTGCAACCGCGCGCAACCGCGCGCTGTGAGCTTGGCGCGAGCGCG 301
QY 1859 AGGTAAAGCTGCTGAGCAGCAGCTGCTGATGAGCAAACTGTATGCGCGCGAGGA 1918
DB 302 GGGCGCGCTGTGTGTGCGCGACCGCGCTTGGCGGGC--CCGTGCGGAGCGCGGTG 359
QY 1919 TCCAGCGGCTGCTGTGAGCAGAGCGCGGTGCAAGCGCAAGCGCAACAGCTTCGATGA 1978
DB 360 ----CGGGCGCGTCCGAGCCGAGAGATCCCGACCGGTCTCGAACCCCGCGGGCCGG 415
QY 1979 TGGCCATTAGACACTTGGGATCTCGCGTATGTCACTTACACTTGGGATCCACAGGGT 2038
DB 416 TCCCGCGCGCGACCGCGCGCAACGCGCGCTACCGCGTTCCTTCACCTCGGTTTCACCGGCC 475
QY 2039 TGGCCAAAGGAGGATGATGATGATCATCGGGGTGCGCTCAACACCATCTGAGCATCAAG 2098
DB 476 GGGCCAAAGGCGCTGTGTGTACCCACGCGGCAATCCGGAACGGATGCGTGAACGCTGC 535
QY 2099 AGCGCTTCAAAATAGGAGCGCGAGACAGAGGTGTGCGCTCTCTGCTGAGCTTGATC 2158
DB 536 AGCGGATGCGCTCGGGGCGCGCGACCGGGTTCTGCAGAAAGACCATGAGGTTGACG 595
QY 2159 TCTCGGTCTATGATGTGTGTGGGATCTGCGCGCGCGGTACGATCTGTGTCCGAGCG 2218
DB 596 CGCGCGCTGAGGATCTTGTGCGCGCGCTGTGCGCGCGGACCGGTGTCTGCCCGCG 655
QY 2219 CGTCAAGCTGCGGATTCGCGCGCATTTGGCAGAGTTGATCGAAGAGAGAGGTGACGG 2278

```

```

DB 656 CGGGGCGGAGCGGAGCCCGGCGCGCTGTGCGCGGTTGCGAGCACGCGGTACCG 715
QY 2279 TGTGAATCTCGTGTCCGCGCGCTGATGCGGATGTCTGTGAGCATTTTGAAGGTCCCGCG 2338
DB 716 TGTCTCAGGTCTGTGCGGTGTGCTGTGCTGTGCTGTGTGAGAGGGTGTGAGAGCGCGC- 774
QY 2339 ATTGCTGCTAGATCTGTGCGCGCTTTTGTGCTGTGAGCGGCACTGTGATCCCGGTGAGCC 2398
DB 775 ----TGGGCTGCTGCGGCTGTGTGTCTGTGCGGGGAGAGGCTTCGACGCGGAGCTG 829
QY 2399 TGTCTGCGAGCTCCAGGCGCATCAGCGCGCGGTGTGATGACCTTGGCGGGGCGCA 2458
DB 830 TGGCGCGGTGTGTGAGGAGAGCGGCGAGACTGTGAGAGTCTGAGAAACCTACGCTGCCA 889
QY 2459 CCGAAGCTGTGATCTGTGCTCATCGGCTACCCGTGAGAGAACCTGACCTATGTGGCGCA 2518
DB 890 CCGAGTGTCTCCATCAGACATCACCGCGAGCTGT--CGACCCCGCGCTCACCGCGGCTC 946
QY 2519 GCATCCCTAGCGCGGTGCGCGCGCAACGAGCTTCCAGCTGTGATGAGGCGCTG 2578
DB 947 CGGTCCCATCTGCGCGCGCGCGCTGCGCGGAGTCCGTCTGTCTCGAGCCGAACGCTG 1006
QY 2579 AACCGCGCGGTGTGAGGTTCCGGGCAACTTACATTTGCGGGTGTGCGCTGAGCACTGG 2638
DB 1007 TCCCGGTGCGGTGTGTGTGCTGCGCGAGCTGTACGCGGGGCGGTTCGAGGTGCGCGG 1066
QY 2639 GCTACTGCGCGATGAGAGAGAGAGCGCGAGAGCTTCTGTGTGAC-----CGG 2689
DB 1067 GCTATGCGCGGCGCGCGCATGTGACCGCGCAACGCTTGTGTGCGCGCATACGCGCACCG 1126
QY 2699 AGACGGGGAGGCGCTTACAAAGACCGGAGATCTGGCGGCTGACCTGCGGATGAGAAACA 2749
DB 1127 AGCCCGGCGCGCTTCTACCGCACCGGTGACAGGTGTGCGGTGCTGCGACCGCACCC 1186
QY 2750 TCGAGTTTATGAGGAGTGAAGCAACAAATCAAGCTTGTGCGATACCGCGTTGAGCTG 2809
DB 1187 TCGAGTACTGTGGCGCGGTGCGACCAAGGTGAGTCAACGCGGTCCGATCGAGCGCG 1246
QY 2810 GGGAAATGAGAGAAAGCTCAAGTGCATCCGACGTACGCAACCGCGGTGATGTGCGCG 2869
DB 1247 CCGAGGTGAGGCGCGCTGCGCGCGCACCGCGCGGTCAACCGCGGTGTGACGCGCT 1306
QY 2870 TCGGGAACGACGCGGGAACAGCTCTTGTGACCTA 2906
DB 1307 ACAGGCGGACGCGCGCGGCAAGCGGCTGCGCGCTA 1343

```

Search completed: October 4, 2003, 12:40:37  
Job time : 1029.11 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 3, 2003, 11:46:26 ; Search time 224.251 Seconds  
(without alignments)  
8331.640 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_11872\_16104  
Perfect score: 4233  
Sequence: 1 atgacgacatcaccgctctc.....aggcgagagagctagctaa 4233

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCOTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfillseq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4233	100.0	68750	3 US-09-335-409-1	Sequence 1, Appli
2	4233	100.0	68750	4 US-09-568-102-1	Sequence 1, Appli
3	4233	100.0	68750	4 US-09-567-969-1	Sequence 1, Appli
4	4233	100.0	68750	4 US-09-568-480-1	Sequence 1, Appli
5	4233	100.0	68750	4 US-09-568-486-1	Sequence 1, Appli
6	4233	100.0	68750	4 US-09-568-472-1	Sequence 1, Appli
7	4233	100.0	68750	4 US-09-567-899-1	Sequence 1, Appli
8	4114.6	97.2	71989	4 US-09-443-501A-2	Sequence 2, Appli
9	520.6	12.3	4563	4 US-09-252-991A-4765	Sequence 4765, Ap
10	491	11.6	2865	4 US-09-252-991A-6775	Sequence 4675, Ap
11	455	10.7	2922	4 US-09-252-991A-4815	Sequence 4815, Ap
12	358.2	8.5	4411529	3 US-09-103-840A-1	Sequence 1, Appli
13	349.2	8.2	4403765	3 US-09-103-840A-2	Sequence 2, Appli
14	348.6	8.2	18442	4 US-09-252-991A-6768	Sequence 4678, Ap
15	340	8.0	2844	4 US-09-252-991A-4764	Sequence 4764, Ap
16	274.4	6.5	3315	4 US-09-252-991A-8892	Sequence 8892, Ap
17	274.4	6.5	7911	4 US-09-252-991A-1882	Sequence 8882, Ap
18	261.4	6.2	4236	4 US-09-252-991A-7057	Sequence 7057, Ap
19	261.4	6.2	10023	4 US-09-252-991A-6997	Sequence 6997, Ap
20	246.6	5.8	6858	4 US-09-252-991A-1219	Sequence 1219, Ap
21	241.6	5.7	77536	4 US-09-410-551B-1	Sequence 1, Appli
22	240.8	5.7	1320	4 US-09-252-991A-4790	Sequence 4790, Ap
23	229.8	5.4	6573	4 US-09-252-991A-9183	Sequence 8888, Ap
24	229.6	5.4	1857	4 US-09-252-991A-8888	Sequence 8888, Ap
25	226.6	5.4	630	4 US-09-252-991A-4792	Sequence 4792, Ap
26	223.2	5.3	2025	4 US-09-252-991A-8891	Sequence 8891, Ap
27	220.2	5.2	7374	4 US-09-252-991A-9100	Sequence 9100, Ap

C	28	218.8	5.2	1689	4 US-09-252-991A-8717	Sequence 8717, Ap
C	29	202	4.8	4242	4 US-09-252-991A-7056	Sequence 7056, Ap
C	30	200.6	4.7	1965	4 US-09-252-991A-9230	Sequence 9230, Ap
C	31	194	4.6	2274	4 US-09-252-991A-6998	Sequence 6998, Ap
C	32	194	4.6	2823	4 US-09-252-991A-7100	Sequence 7100, Ap
C	33	194	4.6	3456	4 US-09-252-991A-7025	Sequence 7025, Ap
C	34	186	4.4	1509	4 US-09-252-991A-63	Sequence 63, Appli
C	35	186	4.4	7527	4 US-09-252-991A-71	Sequence 71, Appli
C	36	182.8	4.3	804	4 US-09-252-991A-4819	Sequence 4819, Ap
C	37	175.4	4.1	954	4 US-09-252-991A-4820	Sequence 4820, Ap
C	38	169.2	4.0	2805	4 US-09-252-991A-9064	Sequence 9064, Ap
C	39	168.2	4.0	2219	3 US-08-510-646B-17	Sequence 17, Appli
C	40	166.2	3.9	1494	4 US-09-252-991A-7049	Sequence 7049, Ap
C	41	163.2	3.9	765	4 US-09-252-991A-4721	Sequence 4721, Ap
C	42	158.2	3.7	1178	3 US-08-861-774E-91	Sequence 91, Appli
C	43	155	3.7	1200	3 US-08-861-774E-85	Sequence 85, Appli
C	44	152.8	3.6	1299	4 US-09-252-991A-7051	Sequence 7051, Ap
C	45	151.4	3.6	3126	4 US-09-252-991A-9066	Sequence 9066, Ap

## ALIGNMENTS

RESULT 1	US-09-335-409-1	Sequence 1, Application US/09335409
Patent No.	6121029	
GENERAL INFORMATION:		
APPLICANT:	Schnupp, Thomas	
APPLICANT:	Ligon, James	
APPLICANT:	Molnar, Istvan	
APPLICANT:	Zirkle, Ross	
APPLICANT:	Cyr, Devon	
APPLICANT:	Goelach, Joern	
TITLE OF INVENTION:	GENES FOR THE BIOSYNTHESIS OF EPOTHILONES	
FILE REFERENCE:	4-30582A	
CURRENT APPLICATION NUMBER:	US/09/335, 409	
CURRENT FILING DATE:	1999-06-17	
NUMBER OF SEQ ID NOS:	30	
SOFTWARE:	Patentin Ver. 2.0	
SEQ ID NO 1		
LENGTH:	68750	
TYPE:	DNA	
ORGANISM:	Sorangium cellulosum	
US-09-335-409-1		
Query Match	100.0%; Score 4233; DB 3; Length 68750;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 4233; Conservative	0; Mismatches	0; Indels
Gaps	0;	
Cy	1	ATGACGATCATCGCTTGAAGAGCTGAGCAGGATATCAAGTGGCGCGAT 60
Db	11872	ATGACGATCATCGCTTGAAGAGCTGAGCAGGATATCAAGTGGCGCGAT 11931
Cy	61	GGGAGGCGCTCGATACAGGCGCCCAAGACCGCTGAACCCGAACCTGCTGCA 120
Db	11932	GGGAGGCGCTCGATACAGGCGCCCAAGACCGCTGAACCCGAACCTGCTGCA 11991
Cy	121	ATTCGACGACCAAAAGCAGATCTTACGATGCTCGTCAGAGACTCCCGCAATTC 180
Db	11992	ATTCGACGACCAAAAGCAGATCTTACGATGCTCGTCAGAGACTCCCGCAATTC 12051
Cy	181	ATTCGACGACCAAAAGCAGATCTTACGATGCTCGTCAGAGACTCCCGCAATTC 240
Db	12052	ATTCGACGACCAAAAGCAGATCTTACGATGCTCGTCAGAGACTCCCGCAATTC 12111
Cy	241	TACTGCTGAGCGCAGACGAGGCTTACGATGCTCCGAGGAGATCCAGCTATCGGAA 300
Db	12112	TACTGCTGAGCGCAGACGAGGCTTACGATGCTCCGAGGAGATCCAGCTATCGGAA 12171
Cy	301	TACGACTGTACGATCTTCGAGCTGCGAGGCTGAGCGGCTTTGCGAAGTGTGCGG 360
Db	12172	TACGACTGTACGATCTTCGAGCTGCGAGGCTGAGCGGCTTTGCGAAGTGTGCGG 12231

QY 361 CCGCAGCATGCTTGGGGCCACAGCTCCCGCATGATGAGGTATCCAGCTTAA 420  
Db 12232 CCGCAGCATGCTTGGGGCCACAGCTCCCGCATGATGAGGTATCCAGCTTAA 12291  
QY 421 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGAGAGCAACGGGAAGG 480  
Db 12292 GTCGACGCCGACATCGAGATCATCGATCTGCGCGGGCTCGACCGAGAGCAACGGGAAGG 12351  
QY 481 AAGCTGTGTGTTGGAGATGCGATGTCGACCCGATCTATGACACGAGCGCCCTCCG 540  
Db 12352 AAGCTGTGTGTTGGAGATGCGATGTCGACCCGATCTATGACACGAGCGCCCTCCG 12411  
QY 541 CTCATACAGTGTGTCGGCTGTGAGTCGACGAGCGCAACCCGCTCTGCTCACTATC 600  
Db 12412 CTCATACAGTGTGTCGGCTGTGAGTCGACGAGCGCAACCCGCTCTGCTCACTATC 12471  
QY 601 GATCTCATTAAGCTTACCTAGGACCTGTCCATCATCTTCAAGGACTGGCTCACTTC 660  
Db 12472 GATCTCATTAAGCTTACCTAGGACCTGTCCATCATCTTCAAGGACTGGCTCACTTC 12531  
QY 661 TACGAAGATCCGAGACCTCTCTCCCTGTCTGAGCTCTCTGACCGGATTAATGATC 720  
Db 12532 TACGAAGATCCGAGACCTCTCTCCCTGTCTGAGCTCTCTGACCGGATTAATGATC 12591  
QY 721 GGGCTGAGTCTTCGCAAGAGTCTGAGGGGATCAACGATCCATGATTAATGAGCGG 780  
Db 12592 GGGCTGAGTCTTCGCAAGAGTCTGAGGGGATCAACGATCCATGATTAATGAGCGG 12651  
QY 781 CGCATCGCGAGCTCCACCTCCGCGAGCTTCGATGAAGGCGATCATCACTTC 840  
Db 12652 CGCATCGCGAGCTCCACCTCCGCGAGCTTCGATGAAGGCGATCATCACTTC 12711  
QY 841 AAGAGATCCGCTTCGAGCAACGAGCAATGCTCCGTCGAGCTCTGGGGTCGATTG 900  
Db 12712 AAGAGATCCGCTTCGAGCAACGAGCAATGCTCCGTCGAGCTCTGGGGTCGATTG 12771  
QY 901 AAGCGCGTGTGCGGGAGCGCGGGTGAACCCGAGCGGGTCACTCTGCTGATTTTC 960  
Db 12772 AAGCGCGTGTGCGGGAGCGCGGGTGAACCCGAGCGGGTCACTCTGCTGATTTTC 12831  
QY 961 GAGGTATCGGGGCTGAGACGCGAGCCCGGGTTACCTCAACATAAGCTCTTCAAC 1020  
Db 12832 GAGGTATCGGGGCTGAGACGCGAGCCCGGGTTACCTCAACATAAGCTCTTCAAC 12891  
QY 1021 CGGCTCCCGTCCATCGCGCGTGAACGATATCACCGGGGACTTCACTGATGCTTC 1080  
Db 12892 CGGCTCCCGTCCATCGCGCGTGAACGATATCACCGGGGACTTCACTGATGCTTC 12951  
QY 1081 CTGGAATGAGACCACTCGGCAAGAGCTTGAACAGCGGCTAAAGGATTTCAAGAG 1140  
Db 12952 CTGGAATGAGACCACTCGGCAAGAGCTTGAACAGCGGCTAAAGGATTTCAAGAG 13011  
QY 1141 CAGCTGTGGGAAGCGATGATCACTGCAAGTAAGCGTATCAGGTCCAGCGAGGCG 1200  
Db 13012 CAGCTGTGGGAAGCGATGATCACTGCAAGTAAGCGTATCAGGTCCAGCGAGGCG 13071  
QY 1201 GCCCGGGTCTTGGGATCAACGAGCGCATTTGTTCCCGTGTGCTCAAGACGCGCTT 1260  
Db 13072 GCCCGGGTCTTGGGATCAACGAGCGCATTTGTTCCCGTGTGCTCAAGACGCGCTT 13131  
QY 1261 AACCAAGATGCTTGTGTCACCTGTTGAGAGGCTGGAACCTCCGATTAACAACAG 1320  
Db 13132 AACCAAGATGCTTGTGTCACCTGTTGAGAGGCTGGAACCTCCGATTAACAACAG 13191  
QY 1321 ACGCAACTCTTCAAGCTGTGATCATCAGCTTACAGAGCAGATGGGACCTGCTC 1380  
Db 13192 ACGCAACTCTTCAAGCTGTGATCATCAGCTTACAGAGCAGATGGGACCTGCTC 13251  
QY 1381 CTCGCGTGGAGCAATGCTGACGAGAGTGTTCGCGCCGACTTCTGAGCAATGCTCGAA 1440  
Db 13252 CTCGCGTGGAGCAATGCTGACGAGAGTGTTCGCGCCGACTTCTGAGCAATGCTCGAA 13311

QY 1441 GGTAGCTGTTTTTCTCCGGGCTCACTAGAGAACCATGGGGTGAACAGGTGCGCTGT 1500  
Db 13312 GGTAGCTGTTTTTCTCCGGGCTCACTAGAGAACCATGGGGTGAACAGGTGCGCTGT 13371  
QY 1501 TCGCTTCGCGCTGCCAGCTAGAAAGCGGGGAGAGCGCAACCGCATCGCTGTG 1560  
Db 13372 TCGCTTCGCGCTGCCAGCTAGAAAGCGGGGAGAGCGCAACCGCATCGCTGTG 13431  
QY 1561 ACGGACATAGCTGACAGGCTGTTCCGGGCGGGGTGAGACAGCTGCCATGACCTC 1620  
Db 13432 ACGGACATAGCTGACAGGCTGTTCCGGGCGGGGTGAGACAGCTGCCATGACCTC 13491  
QY 1621 GCCGTGTGTGGGGGAGAGCGCTCAAGTAAGAGCTTTGGCGCGCTTGGGGGGA 1680  
Db 13492 GCCGTGTGTGGGGGAGAGCGCTCAAGTAAGAGCTTTGGCGCGCTTGGGGGGA 13551  
QY 1681 CTGGCGCGGCTGCGGAGAGGGGACGCGCGCAACATTTGTCGCGTGTGATG 1740  
Db 13552 CTGGCGCGGCTGCGGAGAGGGGACGCGCGCAACATTTGTCGCGTGTGATG 13611  
QY 1741 GAGAAAGCTGGAGAGAGGTTGTCCGGTCTTCGGGTGCTGAGTCAGGGCGGCTAC 1800  
Db 13612 GAGAAAGCTGGAGAGAGGTTGTCCGGTCTTCGGGTGCTGAGTCAGGGCGGCTAC 13671  
QY 1801 GTGCCGATCGAATGCCGACCTTACCGGGAGAGGTATCCACTCTGATCATGTGAG 1860  
Db 13672 GTGCCGATCGAATGCCGACCTTACCGGGAGAGGTATCCACTCTGATCATGTGAG 13731  
QY 1861 GTAAAGCTGTGCTGACAGCAGCATGCTGATGCAAACTGTCACTGCGCGGGATC 1920  
Db 13732 GTAAAGCTGTGCTGACAGCAGCATGCTGATGCAAACTGTCACTGCGCGGGATC 13791  
QY 1921 CAGCGGCTGCTGTGAGCGAGCGCGGCTCGAAGCGCACGCGACACCTTCCATGATG 1980  
Db 13792 CAGCGGCTGCTGTGAGCGAGCGCGGCTCGAAGCGCACGCGACACCTTCCATGATG 13851  
QY 1981 CCATTTGAGACACTTGGATCGCGTATGATCATCTACCTCGGGATTCACAGGGTTG 2040  
Db 13852 CCATTTGAGACACTTGGATCGCGTATGATCATCTACCTCGGGATTCACAGGGTTG 13911  
QY 2041 CCCAAGGGGGTATGATGATCATATCGGGGTGCGGTCAACACCATCTGAGACATCAAGAG 2100  
Db 13912 CCCAAGGGGGTATGATGATCATATCGGGGTGCGGTCAACACCATCTGAGACATCAAGAG 13971  
QY 2101 CGCTTCGAATATGGGCGCGGAGACAGGGTGTCTGCGCTCTCTGCTGAGCTTCAATCTC 2160  
Db 13972 CGCTTCGAATATGGGCGCGGAGACAGGGTGTCTGCGCTCTCTGCTGAGCTTCAATCTC 14031  
QY 2161 TCGGTCTATGATGTGTTGCGGATCTGCGCGGGGCGGTAAGATGCTGTCGAGCGG 2220  
Db 14032 TCGGTCTATGATGTGTTGCGGATCTGCGCGGGGCGGTAAGATGCTGTCGAGCGG 14091  
QY 2221 TCCAAAGCTGCGGATCCGCGCATTTGGCAGATTGATCGAAACGAGAGGTGACGGTG 2280  
Db 14092 TCCAAAGCTGCGGATCCGCGCATTTGGCAGATTGATCGAAACGAGAGGTGACGGTG 14151  
QY 2281 TGGAACTGCTGCTCGCGCTGATGCTGCTGAGACATTTTGAAGGTGCGCCCAT 2340  
Db 14152 TGGAACTGCTGCTCGCGCTGATGCTGCTGAGACATTTTGAAGGTGCGCCCAT 14211  
QY 2341 TCGCTCGTAGAGTCTGCGGCTTTGCTGCTGAGCGGAGCTGATCCCGTGGGCTG 2400  
Db 14212 TCGCTCGTAGAGTCTGCGGCTTTGCTGCTGAGCGGAGCTGATCCCGTGGGCTG 14271  
QY 2401 CTTGGCAGCTTCCAGGACATCAGGCGCGGCTGTGCTGATCAGCTTGGCGGGCCACG 2460  
Db 14272 CTTGGCAGCTTCCAGGACATCAGGCGCGGCTGTGCTGATCAGCTTGGCGGGCCACG 14331  
QY 2461 GAAAGCTGATTTGCTCATCGGGTAAACCCCGTGAAGAAAGTGAACATTCCTGGCGAGC 2520  
Db 14332 GAAAGCTGATTTGCTCATCGGGTAAACCCCGTGAAGAAAGTGAACATTCCTGGCGAGC 14391  
QY 2521 ATCCCTTAGGCGGTCCGCTGCGAACAGAGCTTCAAGTGTGATGAGGCGCTCGAA 2580

```

Db 14392 ATCCCTACGCGCGCTCCCTGGCGAACAGAGCTTCAAGTGTGATGAGCGCTCCGAA 14451
Qy 2581 CGCGCGCGCGCTTGGGTTCCGGGGGCACTTACATTTGGGGGGGGTGGGCTTGGCACTGGAGC 2640
Db 14452 CGCGCGCGCGCTTGGGTTCCGGGGGCACTTACATTTGGGGGGGGTGGGCTTGGCACTGGAGC 14511
Qy 2641 TACTGGCGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
Db 14512 TACTGGCGCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14571
Qy 2701 CGCCTTACAGAGACCGCGGATTTGGGCGGCTACCTGCGCGATGAGAGAGAGAGAGAGAG 2760
Db 14572 CGCCTTACAGAGACCGCGGATTTGGGCGGCTACCTGCGCGATGAGAGAGAGAGAGAGAG 14631
Qy 2761 GGGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Db 14632 GGGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14691
Qy 2821 GAAAGCTCAAGTCCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Db 14692 GAAAGCTCAAGTCCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14751
Qy 2881 GCGCGGAAAGAGCTCTTCTAGCTTATGTGTCTCCGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Db 14752 GCGCGGAAAGAGCTCTTCTAGCTTATGTGTCTCCGAGAGAGAGAGAGAGAGAGAGAGAG 14811
Qy 2941 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Db 14812 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14871
Qy 3001 GACGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Db 14872 GACGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14931
Qy 3061 GACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
Db 14932 GACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14991
Qy 3121 GACGCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
Db 14992 GACGCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15051
Qy 3181 GAGTTGGTCAATTCCTGAGCTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Db 15052 GAGTTGGTCAATTCCTGAGCTGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15111
Qy 3241 AAATTCGTTATCCATCGGCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
Db 15112 AAATTCGTTATCCATCGGCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15171
Qy 3301 TCCGCGCGCATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
Db 15172 TCCGCGCGCATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15231
Qy 3361 TTGCTGAAGAGCTCTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Db 15232 TTGCTGAAGAGCTCTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15291
Qy 3421 GTGTTGATGAAGCGGCGTTCAACTCTGTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Db 15292 GTGTTGATGAAGCGGCGTTCAACTCTGTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 15351
Qy 3481 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3540
Db 15352 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 15411
Qy 3541 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3600
Db 15412 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 15471
Qy 3601 GAAAGGTTGGCGCGGCTTCTGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660

```

```

Db 15472 GAAAGGTTGGCGCGGCTTCTGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15531
Qy 3661 GCGCGGCGCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Db 15532 GCGCGGCGCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15591
Qy 3721 AGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
Db 15592 AGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15651
Qy 3781 CGCGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Db 15652 CGCGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15711
Qy 3841 GATGCGTTGCGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 15712 GATGCGTTGCGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15771
Qy 3901 GATACCTGTCGCGCGGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Db 15772 GATACCTGTCGCGCGGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15831
Qy 3961 CTGCTGCGGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020
Db 15832 CTGCTGCGGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15891
Qy 4021 GTGATCTGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Db 15892 GTGATCTGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15951
Qy 4081 CTGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Db 15952 CTGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16011
Qy 4141 TCCGTTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
Db 16012 TCCGTTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16071
Qy 4201 GAGGTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4233
Db 16072 GAGGTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16104

```

```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 100.0%; Score 4233; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGACGATCAATCAAGCTTCTGAGAGAGCTCGAGCAGGAGATATCAAGCTGGCGCGCAT 60

```



Db 11872 ATGACATCAATAGCTTCTGAACGACTCGAGCACAGGGTATCAAGCTGGGCGCAT 11931  
Qy 61 GGGAGCGCTCCAGATACAGGCCCCAAGAACGCCCCGTAACCCGACCTGCTCGCA 120  
Db 11932 GGGAGCGCTCCAGATACAGGCCCCAAGAACGCCCCGTAACCCGACCTGCTCGCA 11991  
Qy 121 ATCTCGAGACAAAAGACAGATCTGAGATGCTCCGTCAAGATCTCCCCGCAATTC 180  
Db 11992 ATCTCGAGACAAAAGACAGATCTGAGATGCTCCGTCAAGATCTCCCCGCAATTC 12051  
Qy 181 ATCTGCGCGCCCGACGAGCGGCGAGGCTCCGTTCTCTCAAGACATCCAAAGATCC 240  
Db 12052 ATCTGCGCGCCCGACGAGCGGCGAGGCTCCGTTCTCTCAAGACATCCAAAGATCC 12111  
Qy 241 TACTGCTGGGCGGACAGAGCGTTTACGATCCCGAGCGGATCCAGCCTATCGCA 300  
Db 12112 TACTGCTGGGCGGACAGAGCGTTTACGATCCCGAGCGGATCCAGCCTATCGCA 12171  
Qy 301 TACGACTGTAACGATCTCGACGTGCGGAGCTGAGCGCGCTTTGGAAAGTCTCGC 360  
Db 12172 TACGACTGTAACGATCTCGACGTGCGGAGCTGAGCGCGCTTTGGAAAGTCTCGC 12231  
Qy 361 CGGACAGACATGCTTGGGCGCCACAGCTGCCGACATGATGAGGATTCGAGCTTAA 420  
Db 12232 CGGACAGACATGCTTGGGCGCCACAGCTGCCGACATGATGAGGATTCGAGCTTAA 12291  
Qy 421 GTGACCGCCGACATCGATGATCATGATCTGCGGCGCTCGACCGAGACACACGGGAAGC 480  
Db 12292 GTGACCGCCGACATCGATGATCATGATCTGCGGCGCTCGACCGAGACACACGGGAAGC 12351  
Qy 481 AGGCTGTGTCTGTGGAGATGAGATGTGACACCGATCTATGACACCGAGCGCTCCG 540  
Db 12352 AGGCTGTGTCTGTGGAGATGAGATGTGACACCGATCTATGACACCGAGCGCTCCG 12411  
Qy 541 CTCTACACGTCGTCGCGCTCGGCTGACAGCGGCAAAACCGTCTCGTCACTATTC 600  
Db 12412 CTCTACACGTCGTCGCGCTCGGCTGACAGCGGCAAAACCGTCTCGTCACTATTC 12471  
Qy 601 GATCTCATTAACGTTGACCTAGGACGCTGTCCATCATCTTCAAGGACTGGCTCACTTC 660  
Db 12472 GATCTCATTAACGTTGACCTAGGACGCTGTCCATCATCTTCAAGGACTGGCTCACTTC 12531  
Qy 661 TACGAAGATCCGAGACCTCTCTCTGTCTGTGAGACTCTGTACCGCGATTAATGATTC 720  
Db 12532 TACGAAGATCCGAGACCTCTCTCTGTCTGTGAGACTCTGTACCGCGATTAATGATTC 12591  
Qy 721 GCGCTGAGTCTCGAAGAGTCTGAGCGCATCAAGATCGATGATTACTGGAACGG 780  
Db 12592 GCGCTGAGTCTCGAAGAGTCTGAGCGCATCAAGATCGATGATTACTGGAACGG 12651  
Qy 781 CGCATCGCGAGCTCCACCTCCGCGAGCGCTTCCGATGAAGGCGCATCATACCTTC 840  
Db 12652 CGCATCGCGAGCTCCACCTCCGCGAGCGCTTCCGATGAAGGCGCATCATACCTTC 12711  
Qy 841 AAGAGATCCGCTTCCGCGACACAGGACATGCTCCGTCGGAATCTCTGGGGTCAATTG 900  
Db 12712 AAGAGATCCGCTTCCGCGACACAGGACATGCTCCGTCGGAATCTCTGGGGTCAATTG 12771  
Qy 901 AAGCGCGTGTCCGAGGAGCGGCGCTGACCCCGAGCGGCGTCACTCTGCTGATTTTC 960  
Db 12772 AAGCGCGTGTCCGAGGAGCGGCGCTGACCCCGAGCGGCGTCACTCTGCTGATTTTC 12831  
Qy 961 GAGGTATGGGCGCTGAGCGGAGCGGCGCTTACGTCACATTAACGCTTTCAAC 1020  
Db 12832 GAGGTATGGGCGCTGAGCGGAGCGGCGCTTACGTCACATTAACGCTTTCAAC 12891  
Qy 1021 CGGCTCCCGTTCATCCGCGGTGAACGATATCACCGGGGACTTCAAGTGTGATTC 1080  
Db 12892 CGGCTCCCGTTCATCCGCGGTGAACGATATCACCGGGGACTTCAAGTGTGATTC 12951  
Qy 1081 CTGGAATGACACGACTCGGACAAAGAGCTTGAACAGCGGCTAAGGCTATTCAAG 1140

Db 12952 CTGGAATGACACGACTCGGACAAAGGCTTCAAGACCGGCTAAGGCTATTCAAG 13011  
Qy 1141 CAGCTGTGGAAAGCATGATGATCACTGCAAGTAAAGGCTATTCAGGTCAGGAGGCC 1200  
Db 13012 CAGCTGTGGAAAGCATGATGATCACTGCAAGTAAAGGCTATTCAGGTCAGGAGGCC 13071  
Qy 1201 GCGCGGGTCTGGGAGTCAACGAGGCGATTTGTTCCCGTGTGCTCAGGAGCGCTT 1260  
Db 13072 GCGCGGGTCTGGGAGTCAACGAGGCGATTTGTTCCCGTGTGCTCAGGAGCGCTT 13131  
Qy 1261 AACCAAGATGTTGTGTCACTCTGTGACAGGCTCGGAACCTCCGATGACACAGC 1320  
Db 13132 AACCAAGATGTTGTGTCACTCTGTGACAGGCTCGGAACCTCCGATGACACAGC 13191  
Qy 1321 ACGGACTCTCACTCTCTGTGATATCACTCTACAGACAGATGGGAGCTCTCTC 1380  
Db 13192 ACGGACTCTCACTCTCTGTGATATCACTCTACAGACAGATGGGAGCTCTCTC 13251  
Qy 1381 CTGCGGTGGGACATGATGAGAGGTTCCCGCGGACCTTCTGGAAGACATGCTCGA 1440  
Db 13252 CTGCGGTGGGACATGATGAGAGGTTCCCGCGGACCTTCTGGAAGACATGCTCGA 13311  
Qy 1441 GGTACGTCTGTTTCTCCGCGGCTCACTGAGAACATGGGATGAACAGGTGCTGT 1500  
Db 13312 GGTACGTCTGTTTCTCCGCGGCTCACTGAGAACATGGGATGAACAGGTGCTGT 13371  
Qy 1501 TCGCTTCCGCTGCCAGCTAGAACGCGCGGAGCGCAAAACGCAACGAGCTCTCTG 1560  
Db 13372 TCGCTTCCGCTGCCAGCTAGAACGCGCGGAGCGCAAAACGCAACGAGCTCTCTG 13431  
Qy 1561 AGCGACATACGCTGACAGGCGCTGTTCCGCGGCGGCTGAGACAGCTGCCATGAGCTC 1620  
Db 13432 AGCGACATACGCTGACAGGCGCTGTTCCGCGGCGGCTGAGACAGCTGCCATGAGCTC 13491  
Qy 1621 GCGGTGTGTGCGCGCGACAGACGCTCAAGTAAAGCTTTCGCGCTTCCGCGGCA 1680  
Db 13492 GCGGTGTGTGCGCGCGACAGACGCTCAAGTAAAGCTTTCGCGCTTCCGCGGCA 13551  
Qy 1681 CTGCGCGCGGCTGCGGAGAGAGGGGACGCGCGAACATTTGTTGCGGTGTGATG 1740  
Db 13552 CTGCGCGCGGCTGCGGAGAGAGGGGACGCGCGAACATTTGTTGCGGTGTGATG 13611  
Qy 1741 GAGAAAGCTGGAGACAGTGTGCGGCTTCCGCGGTCTGAGTCAAGCGCGGCTTAC 1800  
Db 13612 GAGAAAGCTGGAGACAGTGTGCGGCTTCCGCGGTCTGAGTCAAGCGCGGCTTAC 13671  
Qy 1801 GTGCCGATCGATGCGGACCTACCGGCGAGGCTATCACTACCTCTCGATCATGTAG 1860  
Db 13672 GTGCCGATCGATGCGGACCTACCGGCGAGGCTATCACTACCTCTCGATCATGTAG 13731  
Qy 1861 GTAAAGCTGTGCTGACGACGACATGCTGATGGCAACTGTCACTGGCCGCGGGATTC 1920  
Db 13732 GTAAAGCTGTGCTGACGACGACATGCTGATGGCAACTGTCACTGGCCGCGGGATTC 13791  
Qy 1921 CAGCGGCTGTGTGAGCGAGGCGCGGCTGCAAGGCGACGCGGACCAAGCTTCATG 1980  
Db 13792 CAGCGGCTGTGTGAGCGAGGCGCGGCTGCAAGGCGACGCGGACCAAGCTTCATG 13851  
Qy 1981 CCCATTACAGACCTTCCGATCTCGGATGTGATCTACCTCGGATTCACAGGGTTG 2040  
Db 13852 CCCATTACAGACCTTCCGATCTCGGATGTGATCTACCTCGGATTCACAGGGTTG 13911  
Qy 2041 CCCAAGGGGTATGATGATATGAGGGTGGCGTCAACCATCTGAGCATCAAGAG 2100  
Db 13912 CCCAAGGGGTATGATGATATGAGGGTGGCGTCAACCATCTGAGCATCAAGAG 13971  
Qy 2101 CGCTTGAAATGGGCGCGGAGACAGGGTGTGCGGCTCTCTCGCTGAGCTTCATCTC 2160  
Db 13972 CGCTTGAAATGGGCGCGGAGACAGGGTGTGCGGCTCTCTCGCTGAGCTTCATCTC 14031  
Qy 2161 TCGGTCTATGATGTGTGCGGATCTTGCGGCGGCGGAGTGTGATGTCGAGCGG 2220  
Db 14032 TCGGTCTATGATGTGTGCGGATCTTGCGGCGGCGGAGTGTGATGTCGAGCGG 14091



QY 2221 TCCAGCTGCGGATCCGCGCATTTGGGCAAGTTGATCGAAGAGAGAGTGAAGCTG 2280  
DB 14092 TCCAGGCTGCGGATCCGCGCATTTGGGCAAGTTGATCGAAGAGAGAGTGAAGCTG 14151  
QY 2281 TGGAACTGGTGGCGGCGCTGATGCGGATGCTCGTGAAGCATTTTGAAGTGGTCCCGAT 2340  
DB 14152 TGGAACTGGTGGCGGCGCTGATGCGGATGCTCGTGAAGCATTTTGAAGTGGTCCCGAT 14211  
QY 2341 TGGCGCTAGGTCCTGCGGCTTTCGCTGAGCGGCGAATGGAATCCCGTGGGCGTG 2400  
DB 14212 TGGCGCTAGGTCCTGCGGCTTTCGCTGAGCGGCGAATGGAATCCCGTGGGCGTG 14271  
QY 2401 CTTGGCGAGCTTCAGAGCATAGGCGCGGCTGCTGATGATCAAGCTGGGCGGCGCAC 2460  
DB 14272 CTTGGCGAGCTTCAGAGCATAGGCGCGGCTGCTGATGATCAAGCTGGGCGGCGCAC 14331  
QY 2461 GAAGCTGATCTGGTTCATGCGGTAACCCCGTAGAAGAACTGACCTATCTGGGCGAGC 2520  
DB 14332 GAAGCTGATCTGGTTCATGCGGTAACCCCGTAGAAGAACTGACCTATCTGGGCGAGC 14391  
QY 2521 ATCCCTAGCGCGCTGCGCTGCGCAACAGACCTTCCAGCTGATGATGAGGCGCTCGAA 2580  
DB 14392 ATCCCTAGCGCGCTGCGCTGCGCAACAGACCTTCCAGCTGATGATGAGGCGCTCGAA 14451  
QY 2581 CCGGCGCGGCTGCGGCTGCGGCGCACTTACATTTGGGCGGCTGCGCACTGGGCG 2640  
DB 14452 CCGGCGCGGCTGCGGCTGCGGCGCACTTACATTTGGGCGGCTGCGCACTGGGCG 14511  
QY 2641 TACTGGCGGATGAAAGAGAGACGCGCAAGCTTCTCTGTCACCCGCGAGACCGGGAG 2700  
DB 14512 TACTGGCGGATGAAAGAGAGACGCGCAAGCTTCTCTGTCACCCGCGAGACCGGGAG 14571  
QY 2701 CGCCTCTAAGAACCGGCGATCTGGGCGGCTACCTGCGCGATGAAATGAGATTCATG 2760  
DB 14572 CGCCTCTAAGAACCGGCGATCTGGGCGGCTACCTGCGCGATGAAATGAGATTCATG 14631  
QY 2761 GGGCGTGAAGCAACCAATCAAGCTTGGGCGATGACCGGCTGAGCTGGGGAATTCAG 2820  
DB 14632 GGGCGTGAAGCAACCAATCAAGCTTGGGCGATGACCGGCTGAGCTGGGGAATTCAG 14691  
QY 2821 GAAACGCTCAAGTGCATCCGAACGTACCGGACGCGGATGATTTGCGCGGAGACAC 2880  
DB 14692 GAAACGCTCAAGTGCATCCGAACGTACCGGACGCGGATGATTTGCGCGGAGACAC 14751  
QY 2881 GCGCGAACAAGCTCTTCTAGCTTATGCTGCTGCGGAGGAGACACCGGAGCTGCC 2940  
DB 14752 GCGCGAACAAGCTCTTCTAGCTTATGCTGCTGCGGAGGAGACACCGGAGCTGCC 14811  
QY 2941 GAGCAGACGCGAGCTCAAGACCGAGCGGATCGACGCGAGACACCGCGCGAGACG 3000  
DB 14812 GAGCAGACGCGAGCTCAAGACCGAGCGGATCGACGCGAGACACCGCGCGAGACG 14871  
QY 3001 GACGCTTGAAGCGAGCGGAGAGGATGATCAAGCTCGCTCGACACGATCTCCGAGG 3060  
DB 14872 GACGCTTGAAGCGAGCGGAGAGGATGATCAAGCTCGCTCGACACGATCTCCGAGG 14931  
QY 3061 GACCTGACGGAAGACCGCTGCTGATGATCGGCGGAGATCCCGGAGAGCGGCGGCTG 3120  
DB 14932 GACCTGACGGAAGACCGCTGCTGATGATCGGCGGAGATCCCGGAGAGCGGCGGCTG 14991  
QY 3121 GACGCTAGCGGCGCTGCTGATGATCGGAGAGCTTCTTGAAGAGCCCGATTCGGTTGTT 3180  
DB 14992 GACGCTAGCGGCGCTGCTGATGATCGGAGAGCTTCTTGAAGAGCCCGATTCGGTTGTT 15051  
QY 3181 GAGTTTGTGATTCCTGAGCTGCTTGAAGACGAGCGGAGCGGCGGAGCTTCC 3240  
DB 15052 GAGTTTGTGATTCCTGAGCTGCTTGAAGACGAGCGGAGCGGCGGAGCTTCC 15111  
QY 3241 AAATTCGTTATCCATCGGCGGAGACGTAACCGGCTGCAAACTAGCGGATGTCAAA 3300  
DB 15112 AAATTCGTTATCCATCGGCGGAGACGTAACCGGCTGCAAACTAGCGGATGTCAAA 15171

QY 3301 TCCGCGCGATCGAGGCGTGGAGCAGAGGCTTCTATTATTATCAACCCGTTGAGACCGT 3360  
DB 15172 TCCGCGCGATCGAGGCGTGGAGCAGAGGCTTCTATTATTATCAACCCGTTGAGACCGT 15231  
QY 3361 TTGCTGAAGCTCTCCGATCAGCGGATCGAGGCGGAGCGGACGTTCCGCAAACTTCAC 3420  
DB 15232 TTGCTGAAGCTCTCCGATCAGCGGATCGAGGCGGAGCGGACGTTCCGCAAACTTCAC 15291  
QY 3421 GTGTTTCATGAAGCGGCGTTCAACCTCTGTTGCGGCGAGGATCGAGCGCATGAGTCG 3480  
DB 15292 GTGTTTCATGAAGCGGCGTTCAACCTCTGTTGCGGCGAGGATCGAGCGCATGAGTCG 15351  
QY 3481 CTGATGATGATGTCGTGCGCGAATTTTTCCTGCTGAGGCGGAGATATATGCGGAGCTC 3540  
DB 15352 CTGATGATGATGTCGTGCGCGAATTTTTCCTGCTGAGGCGGAGATATATGCGGAGCTC 15411  
QY 3541 CTGATGAGACAGGCGCTTCTGCAACATCGGCGTCTGCTCGGCTGGGCAATTCATTTT 3600  
DB 15412 CTGATGAGACAGGCGCTTCTGCAACATCGGCGTCTGCTCGGCTGGGCAATTCATTTT 15471  
QY 3601 GAACAGGTTGGCGGCTTCTGACCTGCAATTCGAGCGTTTACGTCAGCGGAGTCG 3660  
DB 15472 GAACAGGTTGGCGGCTTCTGACCTGCAATTCGAGCGTTTACGTCAGCGGAGTCG 15531  
QY 3661 GCGCGCGGCTGAGACCGCGCGAGTTCAGGCTGTACGCTCGGTCAGGATTCCTCACCG 3720  
DB 15532 GCGCGCGGCTGAGACCGCGCGAGTTCAGGCTGTACGCTCGGTCAGGATTCCTCACCG 15591  
QY 3721 AGCGCGCGACGACGCGGCGGCTTCCGCGCGGAGACGACCTTGGCGATTCGTT 3780  
DB 15592 AGCGCGCGACGACGCGGCGGCTTCCGCGCGGAGACGACCTTGGCGATTCGTT 15651  
QY 3781 CGGACCTTCTTGAAGACCAACTACCGGAGTACATGAGGCTTACGATCTTGGAGGCTC 3840  
DB 15652 CGGACCTTCTTGAAGACCAACTACCGGAGTACATGAGGCTTACGATCTTGGAGGCTC 15711  
QY 3841 GATGCTTCCGCTGACGCTCCACCGGAGTTCATGAGGCTTCCGCGAGCGAG 3900  
DB 15712 GATGCTTCCGCTGACGCTCCACCGGAGTTCATGAGGCTTCCGCGAGCGAG 15771  
QY 3901 GATACCTGCTGCGCGGCGATTCGGGCGACACGCGGCGACCGGAGCTTGGAGGATC 3960  
DB 15772 GATACCTGCTGCGCGGCGATTCGGGCGACACGCGGCGACCGGAGCTTGGAGGATC 15831  
QY 3961 CTGCTGCGGCTGACGCGGAGGCTGCGGCTGAGAGGCTGCGGCTCAGCAGAGCTC 4020  
DB 15832 CTGCTGCGGCTGACGCGGAGGCTGCGGCTGAGAGGCTGCGGCTCAGCAGAGCTC 15891  
QY 4021 GTGATCTTGTGAGCATCGATTCAATCGTTGCAATGAGAGCTTGTGAGAGG 4080  
DB 15892 GTGATCTTGTGAGCATCGATTCAATCGTTGCAATGAGAGCTTGTGAGAGG 15951  
QY 4081 CTGATAGGAGATCGCATCAACGAGTTGTTCAAGTACCGGAACCTCGGCTCGTGCG 4140  
DB 15952 CTGATAGGAGATCGCATCAACGAGTTGTTCAAGTACCGGAACCTCGGCTCGTGCG 16011  
QY 4141 TCCGCTTGGCGCGAGACTGAGAGATCTATACAGGCGCGGAATTCAGAGAGCTG 4200  
DB 16012 TCCGCTTGGCGCGAGACTGAGAGATCTATACAGGCGCGGAATTCAGAGAGCTG 16071  
QY 4201 GAGGTTGCGCGAGAGCGAGAGAGCTAGCTAA 4233  
DB 16072 GAGGTTGCGCGAGAGCGAGAGAGCTAGCTAA 16104

RESULT 3  
US-09-567-969-1  
; Sequence 1, Application US/09567969  
; Patent No. 635457  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Deyon  
APPLICANT: Goetlich, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/567,969  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 68750  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-09-567-969-1

Query Match 100.0%; Score 4233; DB 4; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGATCAATCAGCTTCTGAAAGAGCTTGAGCAGCAGGATTCAGCTGCGCGCAT 60  
DB 11872 ATGACGATCAATCAGCTTCTGAAAGAGCTTGAGCAGCAGGATTCAGCTGCGCGCAT 11931

QY 61 GGGAGAGCGCTCCAGATACAGGCCCCCAAGAAAGCCCTGAAACCCGAACTGCTGCTCGA 120  
DB 11932 GGGAGAGCGCTCCAGATACAGGCCCCCAAGAAAGCCCTGAAACCCGAACTGCTGCTCGA 11991

QY 121 ATCTCGAGACAAAGACAGATCTGACGATCTCGTACAGATCTCCCGGAGATCC 180  
DB 11992 ATCTCGAGACAAAGACAGATCTGACGATCTCGTACAGATCTCCCGGAGATCC 12051

QY 181 ATCTGCCCCCGCCAGCCGAGCGGACAGCTCGTTTCTCTCCAGACATCCAAAGATCC 240  
DB 12052 ATCTGCCCCCGCCAGCCGAGCGGACAGCTCGTTTCTCTCCAGACATCCAAAGATCC 12111

QY 241 TACTGCTGCGCGGACAGAGCGTTTACGATCCCGAGCGGATTCACAGCCTATCCGCGAA 300  
DB 12112 TACTGCTGCGCGGACAGAGCGTTTACGATCCCGAGCGGATTCACAGCCTATCCGCGAA 12171

QY 301 TAGCACTGACGATCTCGACGTGCGGAGCTGAGCGCGCTTTCGAAAGTCTGCGCG 360  
DB 12172 TAGCACTGACGATCTCGACGTGCGGAGCTGAGCGCGCTTTCGAAAGTCTGCGCG 12231

QY 361 CGGACGACATGCTTGGGCGCCACACGCTGCGCGGACATGATGAGGCTTAA 420  
DB 12232 CGGACGACATGCTTGGGCGCCACACGCTGCGCGGACATGATGAGGCTTAA 12291

QY 421 GTGAGCGCGGACATGAGATCATGATCTGCGGCGGCTGACCGGAGCAACCGGAAAGCG 480  
DB 12292 GTGAGCGCGGACATGAGATCATGATCTGCGGCGGCTGACCGGAGCAACCGGAAAGCG 12351

QY 481 AGGCTGCTGCTTGGAGATGCGATGCGCACCGATCTATGACACCGACCGCTCGG 540  
DB 12352 AGGCTGCTGCTTGGAGATGCGATGCGCACCGATCTATGACACCGACCGCTCGG 12411

QY 541 CTCTATCAGCTGCTGCGCGCTGCGGCTGAGCAGCGGCAACCCGCTCTGCTCACTATC 600  
DB 12412 CTCTATCAGCTGCTGCGCGCTGCGGCTGAGCAGCGGCAACCCGCTCTGCTCACTATC 12471

QY 601 GATCTCATTAACCTTAACTGAGGACGCTGTCATATCATCTTAAAGACCTGCGTCACTTC 660  
DB 12472 GATCTCATTAACCTTAACTGAGGACGCTGTCATATCATCTTAAAGACCTGCGTCACTTC 12531

QY 661 TAGCAAGATCCGAGACCTTCTCCCTGCTGAGAGCTCGTACCGCGATTAATGATCT 720  
DB 12532 TAGCAAGATCCGAGACCTTCTCCCTGCTGAGAGCTCGTACCGCGATTAATGATCT 12591

QY 721 GCGCTGAGCTCTGCAAGAAAGTCTGAGCGCATCAACGATGATGATTACTGAAACCG 780  
DB 12592 GCGCTGAGCTCTGCAAGAAAGTCTGAGCGCATCAACGATGATGATTACTGAAACCG 12651

QY 781 GCGATCGCGAGCTCCACCTCCCGGACGCTTCCGATGAAGCGCATCATCACTCCG 840  
DB 12652 GCGATCGCGAGCTCCACCTCCCGGAGCTTCCGATGAAGCGCATCATCACTCCG 12711

QY 841 AAGGATCCGCTTCCGACACAGAGATGCTGCGCTGAGATCTCTGAGGCTGATG 900  
DB 12712 AAGGATCCGCTTCCGACACAGAGATGCTGCGCTGAGATCTCTGAGGCTGATG 12771

QY 901 AAGCGGCTGCTGAGGAGCGCGGCTGACCCCGAGCGCTCATCTGCTGATTTTCC 960  
DB 12772 AAGCGGCTGCTGAGGAGCGCGGCTGACCCCGAGCGCTCATCTGCTGATTTTCC 12831

QY 961 GAGTGATCGGCGGCTGAGCGCGGAGCGCGGCTTACGCTCAACATACGCTTCAAC 1020  
DB 12832 GAGTGATCGGCGGCTGAGCGCGGAGCGCGGCTTACGCTCAACATACGCTTCAAC 12891

QY 1021 CGGCTCCCGCTCATCCGCGCGCTGACAGATATCACCGGAGACTTACAGTGTGCTTC 1080  
DB 12892 CGGCTCCCGCTCATCCGCGCGCTGACAGATATCACCGGAGACTTACAGTGTGCTTC 12951

QY 1081 CTGACATTCGACACCACTCGGACAAAGCTTCCGAAACCGCGCTTAAGCTATTCAGAG 1140  
DB 12952 CTGACATTCGACACCACTCGGACAAAGCTTCCGAAACCGCGCTTAAGCTATTCAGAG 13011

QY 1141 CAGCTGAGGAAAGCATGATGATCATCGACGATGACGCTTATGAGTCCAGCGAGGCG 1200  
DB 13012 CAGCTGAGGAAAGCATGATGATCATCGACGATGACGCTTATGAGTCCAGCGAGGCG 13071

QY 1201 GCCCGGCTCTGAGGATCCAAAGAGCGCATTTTCCCGTGTGCTGACAGAGCGCTT 1260  
DB 13072 GCCCGGCTCTGAGGATCCAAAGAGCGCATTTTCCCGTGTGCTGACAGAGCGCTT 13131

QY 1261 AACCGAAGTGTGTTGTCTACCTCTGTCGAGAGGCTCGAACTCCGCTGATCAACAGC 1320  
DB 13132 AACCGAAGTGTGTTGTCTACCTCTGTCGAGAGGCTCGAACTCCGCTGATCAACAGC 13191

QY 1321 AGCAGACTCTCAGCTGCTGATGATCATGCTCTACAGACAGATGAGGAGCTGCTG 1380  
DB 13192 AGCAGACTCTCAGCTGCTGATGATCATGCTCTACAGACAGATGAGGAGCTGCTG 13251

QY 1381 CTGCGGTGGACATGCTGACGAGATGTTCCGCGCCAGCTTCTGACGACATGCTCGAA 1440  
DB 13252 CTGCGGTGGACATGCTGACGAGATGTTCCGCGCCAGCTTCTGACGACATGCTCGAA 13311

QY 1441 GGTACGTGTTTTTCTCCGCGGCTCACTGAGAACATGAGGAGTGAACAGGTGCGCTG 1500  
DB 13312 GGTACGTGTTTTTCTCCGCGGCTCACTGAGAACATGAGGAGTGAACAGGTGCGCTG 13371

QY 1501 TCGCTTCCGCTGCGCAGCTAAGACGCGGAGCGGCAACGCGACCAACCGCTGCTG 1560  
DB 13372 TCGCTTCCGCTGCGCAGCTAAGACGCGGAGCGGCAACGCGACCAACCGCTGCTG 13431

QY 1561 AGCGACATACGCTGACAGGCTGTTCCGCGCGGCTGAGCAGCTGCGCATGACCTC 1620  
DB 13432 AGCGACATACGCTGACAGGCTGTTCCGCGCGGCTGAGCAGCTGCGCATGACCTC 13491

QY 1621 GCCGTGTGTGCGCGCGCAAGACCTCACTGACAAAGCTTTCGCGCGCTGCGGCGGA 1680  
DB 13492 GCCGTGTGTGCGCGCGCAAGACCTCACTGACAAAGCTTTCGCGCGCTGCGGCGGA 13551

QY 1681 CTGAGCGCGGCTCGCGAGAGGAGGCGCGCGCAACATTTGCTGCGGTGATG 1740  
DB 13552 CTGAGCGCGGCTCGCGAGAGGAGGCGCGCGCAACATTTGCTGCGGTGATG 13611

QY 1741 GAGAAAGCTGAGAGCAAGTTGTGCGGTTCTGCGGTGCTGAGTCAAGCGCGCTTAC 1800  
DB 13612 GAGAAAGCTGAGAGCAAGTTGTGCGGTTCTGCGGTGCTGAGTCAAGCGCGCTTAC 13671

QY 1801 GTGCGGATGATGCGGACCTTACCGCGAGCGCTTACCTACCTCGATATGATGATG 1860  
DB 13672 GTGCGGATGATGCGGACCTTACCGCGAGCGCTTACCTACCTCGATATGATGATG 13731

QY 1861 GTAAAGCTGTGCTGACGAGCATGAGTGAATGAGCAACTGCTCATGCGCGCGGAGATC 1920

Db 13732 GTAAAGCTCTGTGACACGACCATGCTGATGGCAAACTGTCAATGACCGCGGGATC 13791  
Qy 1991 CAGGGGCTGCTGTGAGGAGGCGGCGTCCGAAAGCGACGCGACACAGCTCCGATGATG 1980  
Db 13792 CAGCGGCTGCTGTGAGGAGGCGGCGTCCGAAAGCGACGCGACACAGCTCCGATGATG 13851  
Qy 1981 CCCATTGACAGACCTTCCGATCTCCGATATGCTCATCTACACTCCGAGATCCACAGGTTG 2040  
Db 13852 CCCTTACAGACACTTCCGATCTCCGATATGCTCATCTACACTCCGAGATCCACAGGTTG 13911  
Qy 2041 CCCAAGGGGGTATGATGATCATCTGAGGATGCTGCAACCATCTTGACATCAACGAG 2100  
Db 13912 CCCAAGGGGGTATGATGATCATCTGAGGATGCTGCAACCATCTTGACATCAACGAG 13971  
Qy 2101 CGCTTCGAATGAGGCGCGGAGACAGGGGCTGAGGCTCTCTCCGCTGACCTTCGATCTC 2160  
Db 13972 CGCTTCGAATGAGGCGCGGAGACAGGGGCTGAGGCTCTCTCCGCTGACCTTCGATCTC 14031  
Qy 2161 TCGCTCTATGATGATGTTGGGATCTCTGGCGGCGGCGGATCGATGCTGGTCCGACGCG 2220  
Db 14032 TCGCTCTATGATGATGTTGGGATCTCTGGCGGCGGCGGATCGATGCTGGTCCGACGCG 14091  
Qy 2221 TCCAAAGCTGCGGATCCGCGCATTTGGGCGAGATTGATCGAAAGAGAGTGAACGCGTG 2280  
Db 14092 TCCAAAGCTGCGGATCCGCGCATTTGGGCGAGATTGATCGAAAGAGAGTGAACGCGTG 14151  
Qy 2281 TGGAACTGGTCCCGGCGCTGATGGGATGCTCTGTCGAGCATTTTGAAGGTCGCCCGAT 2340  
Db 14152 TGGAACTGGTCCCGGCGCTGATGGGATGCTCTGTCGAGCATTTTGAAGGTCGCCCGAT 14211  
Qy 2341 TCGCTCGTAGAGCTCTGCGGCTTTCGCTGTCGAGCGGCGACTGGATCCCGTGGGCGCTG 2400  
Db 14212 TCGCTCGTAGAGCTCTGCGGCTTTCGCTGTCGAGCGGCGACTGGATCCCGTGGGCGCTG 14271  
Qy 2401 CTGCGGAGCTTCAGGCGCATCAGGCCCGGCGTGTGCTGATCAGCTTGGCGGGGCCAAC 2460  
Db 14272 CTGCGGAGCTTCAGGCGCATCAGGCCCGGCGTGTGCTGATCAGCTTGGCGGGGCCAAC 14311  
Qy 2461 GAAAGGTGATCTGTGTCATCGGATACCCCGTGAAGAAAGTGCATCTCTGCGGCGAGC 2520  
Db 14312 GAAAGGTGATCTGTGTCATCGGATACCCCGTGAAGAAAGTGCATCTCTGCGGCGAGC 14391  
Qy 2521 ATCCCTACGCGCGCTCCGCTGCGCAACGAGAGCTTCCAGTCTCGATGAGGCGCTCGAA 2580  
Db 14392 ATCCCTACGCGCGCTCCGCTGCGCAACGAGAGCTTCCAGTCTCGATGAGGCGCTCGAA 14451  
Qy 2581 CCGCGCGCGCTGCGGTTCCGCGGCACTCTACATTTGGCGGGTGGGCTGGCACTGGGC 2640  
Db 14452 CCGCGCGCGCTGCGGTTCCGCGGCACTCTACATTTGGCGGGTGGGCTGGCACTGGGC 14511  
Qy 2641 TACTGGCCCGATGAGAGAGAGCGCGCAAGAGCTTCTCTGTGACACCCGAGACCGGGAG 2700  
Db 14512 TACTGGCCCGATGAGAGAGAGCGCGCAAGAGCTTCTCTGTGACACCCGAGACCGGGAG 14571  
Qy 2701 CGCCTCTACAGACCGGCGATCTGGGCGGCTACCTGCCGATGAAACATCGAGTTCTAG 2760  
Db 14572 CGCCTCTACAGACCGGCGATCTGGGCGGCTACCTGCCGATGAAACATCGAGTTCTAG 14611  
Qy 2761 GGGCGTGAAGACCAACCAATCAAGCTTCCGCGATACCGGTTGAGCTCGGGGAAATCGAG 2820  
Db 14612 GGGCGTGAAGACCAACCAATCAAGCTTCCGCGATACCGGTTGAGCTCGGGGAAATCGAG 14691  
Qy 2821 GAAAGCTCAAGTCCGATCCGAAAGTACGCGACGCGGTGATGTTGTCCTCGGAGAACAC 2880  
Db 14692 GAAAGCTCAAGTCCGATCCGAAAGTACGCGACGCGGTGATGTTGTCCTCGGAGAACAC 14751  
Qy 2881 GCGGCGAACAGAGCTCTTCTAGCTATGATGCTCCGAGAGGACACGAGAGACGCGCTGCC 2940  
Db 14752 GCGGCGAACAGAGCTCTTCTAGCTATGATGCTCCGAGAGGACACGAGAGACGCGCTGCC 14811  
Qy 2941 GAGCAGAGCGGAGCTTCAAGACCGAGCGGATCGACGCGAGAGCACGCGCGCGAAGCG 3000

Db 14812 GAGCAGAGCGGAGCGCTTCAAGACCGAGCGGATCGACCGGAGAGCACGCGCGGAAACG 14871  
Qy 3001 GACGCGCTTACAGCGACGCGGAGAGGATGATTAAGTCTGCTCGACACGAGCTCCGAGG 3060  
Db 14872 GACGCGCTTACAGCGACGCGGAGAGGATGATTAAGTCTGCTCGACACGAGCTCCGAGG 14931  
Qy 3061 GACCTGACGGAAGGCCGCTGCTGATCTGACCGGGCAGATTCGCGGGAGCGGGGCTG 3120  
Db 14932 GACCTGACGGAAGGCCGCTGCTGATCTGACCGGGCAGATTCGCGGGAGCGGGGCTG 14991  
Qy 3121 GACCTGACGCGCGCTCCGCTGAGGCTCGAACGTTCTCTTGAAGCCCGATTCCTTTGTT 3180  
Db 14992 GACCTGACGCGCGCTCCGCTGAGGCTCGAACGTTCTCTTGAAGCCCGATTCCTTTGTT 15051  
Qy 3181 GAGTTGGTGCATTCCTGAGCTGCTGAGCGCGTGAAGCGCGGACCGGACGCGGCGACCTTCC 3240  
Db 15052 GAGTTGGTGCATTCCTGAGCTGCTGAGCGCGTGAAGCGCGGACCGGACGCGGCGACCTTCC 15111  
Qy 3241 AAATTCGTTATCCATCCGCGGAGACAGTACCCGCTGCAAACTACGCGTATGTCAAA 3300  
Db 15112 AAATTCGTTATCCATCCGCGGAGACAGTACCCGCTGCAAACTACGCGTATGTCAAA 15171  
Qy 3301 TCCGCGCGCATCGAGGCGGTGAGAGAGGCTTCTATTATTCACCCGTTGAGACCGCT 3360  
Db 15172 TCCGCGCGCATCGAGGCGGTGAGAGAGGCTTCTATTATTCACCCGTTGAGACCGCT 15231  
Qy 3361 TTGCGAAGCTCTCGATCAGCGGATCGAGGCGGAGGCGAGCGTTCCGCAAACTTCGAC 3420  
Db 15232 TTGCGAAGCTCTCGATCAGCGGATCGAGGCGGAGGCGAGCGTTCCGCAAACTTCGAC 15291  
Qy 3421 GTGTTCATGAAAGCGGCGTTCAACTCTCTGTTCTGGCAGAGATCGACCGCATCGAGTCG 3480  
Db 15292 GTGTTCATGAAAGCGGCGTTCAACTCTCTGTTCTGGCAGAGATCGACCGCATCGAGTCG 15351  
Qy 3481 CTGTATGATGCTGTGTCGCGAGAAATTTTCTCTGTCGAGCGCGGATATATGCGCAGCTC 3540  
Db 15352 CTGTATGATGCTGTGTCGCGAGAAATTTTCTCTGTCGAGCGCGGATATATGCGCAGCTC 15411  
Qy 3541 CTGATGAGAGAGGCGCTCTCGCAACATCGGCGGCTGTCGCGTGGGCGCAATTCAATTTT 3600  
Db 15412 CTGATGAGAGAGGCGCTCTCGCAACATCGGCGGCTGTCGCGTGGGCGCAATTCAATTTT 15471  
Qy 3601 GAAAGGTTCCGCGCGGTTCTCGACTCGACATTCGACGTTTACGTGACAGCGCATGCTG 3660  
Db 15472 GAAAGGTTCCGCGCGGTTCTCGACTCGACATTCGACGTTTACGTGACAGCGCATGCTG 15531  
Qy 3661 GCGCGGCGGATGACCCGCGGAGTTCCAGGCTGTGACGCTCGGTCAGGATTCCTCACCG 3720  
Db 15532 GCGCGGCGGATGACCCGCGGAGTTCCAGGCTGTGACGCTCGGTCAGGATTCCTCACCG 15591  
Qy 3721 AGGCGCGCACGAGCGCGGCGCCCTCCGCGCGCGGAGCACTTGCGCGATATGCTT 3780  
Db 15592 AGGCGCGCACGAGCGCGGCGCCCTCCGCGCGCGGAGCACTTGCGCGATATGCTT 15651  
Qy 3781 CGGCACTTCTTGAGGACCAAACTACCCGAGTACATGAGCTTACAGTCTTGTGAGGCTC 3840  
Db 15652 CGGCACTTCTTGAGGACCAAACTACCCGAGTACATGAGCTTACAGTCTTGTGAGGCTC 15711  
Qy 3841 GATGCTTCCGCTGACGTTCCACCGGCAAGGTCATGTAAGCCCTGCGCGAGCGGAG 3900  
Db 15712 GATGCTTCCGCTGACGTTCCACCGGCAAGGTCATGTAAGCCCTGCGCGAGCGGAG 15771  
Qy 3901 GATACCTGTCGCGCGGCACTTCCGCGGACACCGGCGCACGAGGAGCTTGAAGAGATC 3960  
Db 15772 GATACCTGTCGCGCGGCACTTCCGCGGACACCGGCGCACGAGGAGCTTGAAGAGATC 15831  
Qy 3961 CTGCTGCGGCTCTGACGAGAGTCTCGGCTGAGAGTGTCTCGGCTTCCAGACAGCTTC 4020  
Db 15832 CTGCTGCGGCTCTGACGAGAGTGTCTCGGCTGAGAGTGTCTCGGCTTCCAGACAGCTTC 15891  
Qy 4021 GTCGATCTTGGTGGACATCGATTCAATGTTGTCATGAGAGGCTTGTGCAAGAGG 4080  
Db 15892 GTCGATCTTGGTGGACATCGATTCAATGTTGTCATGAGAGGCTTGTGCAAGAGG 15951

QY 4081 CTGGATAGGAGATCGCCATCAACCGAGTTGTTCCAGTACCCGAACCTCGGCTGGTGGG 4140  
DB 15952 CTGGATAGGAGATCGCCATCAACCGAGTTGTTCCAGTACCCGAACCTCGGCTGGTGGG 16011  
QY 4141 TCCGTTTCCGCGGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGACGACCGAGT 4200  
DB 16012 TCCGTTTCCGCGGAGACTCGAGAGATCTAGATCAGCGGCCGAACATGACGACCGAGT 16071  
QY 4201 GAGTTTCGCGCAAGGCGAGAGAGCTAGCTAA 4233  
DB 16072 GAGTTTCGCGCAAGGCGAGAGAGCTAGCTAA 16104

RESULT 4  
US-09-568-480-1  
Sequence 1, Application US/09568480  
Patent No. 6355458  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Rose  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,480  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 68750  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-09-568-480-1

Query Match 100.0%; Score 4233; DB 4; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACATCAATAGTCTCGAAGCGATCTGACACAGGCTTCAAGCTGGGCGCGAT 60  
DB 11872 ATGACATCAATAGTCTCGAAGCGATCTGACACAGGCTTCAAGCTGGGCGCGAT 11931  
QY 61 GGGAGAGCGCTCGAGATACAGGCGGCCCAAGACGCGCTGAAACCGAAGCTGCTGCGA 120  
DB 11932 GGGAGAGCGCTCGAGATACAGGCGGCCCAAGACGCGCTGAAACCGAAGCTGCTGCGA 11991  
QY 121 ATCTCCGAGCAAAAGACAGATCTGACGATCTCCGTCAGAGACTCCCGCGAATCC 180  
DB 11992 ATCTCCGAGCAAAAGACAGATCTGACGATCTCCGTCAGAGACTCCCGCGAATCC 12051  
QY 181 ATGTCGCCGCCCGCCAGCGGCGGCGAGCTCCGTTCTCTCAGACATCCAAAGATCC 240  
DB 12052 ATGTCGCCGCCCGCCAGCGGCGGCGAGCTCCGTTCTCTCAGACATCCAAAGATCC 12111  
QY 241 TACTGCTGGGCGCGAGCAGGAGCGTTTACGATCCCGAGCGGATTCACGCTTACCGCA 300  
DB 12112 TACTGCTGGGCGCGAGCAGGAGCGTTTACGATCCCGAGCGGATTCACGCTTACCGCA 12171  
QY 301 TACGACTGTACGATCTCGACGTGCGAGGCTGAGCGCGGCTTTGGAAAAGTGTGCGG 360  
DB 12172 TACGACTGTACGATCTCGACGTGCGAGGCTGAGCGCGGCTTTGGAAAAGTGTGCGG 12231  
QY 361 CGGCAAGACATGTTTGGGCGGCAAGCTGCGCGCGGATGATGAGGTGATTCGAGCTTAA 420  
DB 12232 CGGCAAGACATGTTTGGGCGGCAAGCTGCGCGCGGATGATGAGGTGATTCGAGCTTAA 12291  
QY 421 GTTCAGCGCGACATCGAGATCATGATCTTGGCGGCGCTCGACCGGAGACACGGAAGCG 480

DB 12292 GTTCAGCGCGACATCGAGATCATGATCTTGGCGGCGCTCGACCGGAGACACGGAAGCG 12351  
QY 481 AGGCTGTGTGTTTGGGAGATTCGATGTGCAACCGCATCTATGACACGAGGCGCTCCG 540  
DB 12352 AGGCTGTGTGTTTGGGAGATTCGATGTGCAACCGCATCTATGACACGAGGCGCTCCG 12411  
QY 541 CTCTATCAAGTGTGCGCGGTTGCGGCTGAGCAGGCGGCAAAACCGTCTGCTGCTAGTATC 600  
DB 12412 CTCTATCAAGTGTGCGCGGTTGCGGCTGAGCAGGCGGCAAAACCGTCTGCTGCTAGTATC 12471  
QY 601 GATCTATTAAAGTTGACCTTAGAGAGCTGTTCATCATCTTCAAGAGACTGCTCAGCTTC 660  
DB 12472 GATCTATTAAAGTTGACCTTAGAGAGCTGTTCATCATCTTCAAGAGACTGCTCAGCTTC 12531  
QY 661 TACGAAGATCCGGAAGCTCTCTCCCTGTCTCTGAGAGCTCTGTTACCGGATTAATGATCTC 720  
DB 12532 TACGAAGATCCGGAAGCTCTCTCCCTGTCTCTGAGAGCTCTGTTACCGGATTAATGATCTC 12591  
QY 721 GCGCTGAGGCTCGCAAGAAAGTCTGAGGCGCATCAACGATCGATGATTAATGAGAGCGG 780  
DB 12592 GCGCTGAGGCTCGCAAGAAAGTCTGAGGCGCATCAACGATCGATGATTAATGAGAGCGG 12651  
QY 781 CGCATCGCGAGCTCCACCTCCGCGGAGCGCTTCGATGAGAGCGGATCTACCTTG 840  
DB 12652 CGCATCGCGAGCTCCACCTCCGCGGAGCGCTTCGATGAGAGCGGATCTACCTTG 12711  
QY 841 AAGAGATCCGCTTCGCGGACACCGAGCAATGGCTCCGTCGAGCTCCCGGGGTGATTTG 900  
DB 12712 AAGAGATCCGCTTCGCGGACACCGAGCAATGGCTCCGTCGAGCTCCCGGGGTGATTTG 12771  
QY 901 AAGCGCGGTGTGGGAGAGCGGAGCTGACCCCGAGCGGCGTATCTGCGCTCATTTTTC 960  
DB 12772 AAGCGCGGTGTGGGAGAGCGGAGCTGACCCCGAGCGGCGTATCTGCGCTCATTTTTC 12831  
QY 961 GAGGTATCGGCGCTGAGCGGAGCGGCGGCTTACGCTCAATTAACGCTCTTCAAC 1020  
DB 12832 GAGGTATCGGCGCTGAGCGGAGCGGCGGCTTACGCTCAATTAACGCTCTTCAAC 12891  
QY 1021 CGGCTCCCGCTCATCTCGGCGCTGAAACGATATCAACCGGAGATTTCAACGCTGCTTC 1080  
DB 12892 CGGCTCCCGCTCATCTCGGCGCTGAAACGATATCAACCGGAGATTTCAACGCTGCTTC 12951  
QY 1081 CTGACATGACACCACTGCGGAGCAAGAGCTTCAAGACGCGGCTTAAGCGTATTAAGAG 1140  
DB 12952 CTGACATGACACCACTGCGGAGCAAGAGCTTCAAGACGCGGCTTAAGCGTATTAAGAG 13011  
QY 1141 CAGCTGTGGGAAGCATGATGATCACTGCGAGCTTAACGCGGTATCGAGGCTCCAGAGAGCC 1200  
DB 13012 CAGCTGTGGGAAGCATGATGATCACTGCGAGCTTAACGCGGTATCGAGGCTCCAGAGAGCC 13071  
QY 1201 GCGCGGCTCTGGGATCCAAAGAGCGGATTTCCCGGTGCTGTCAGAGCGGCTT 1260  
DB 13072 GCGCGGCTCTGGGATCCAAAGAGCGGATTTCCCGGTGCTGTCAGAGCGGCTT 13131  
QY 1261 AACCAAGATGTTGTTGCTCACTGTTGACAGAGCTCGGAATCCGCTGTATACACAGC 1320  
DB 13132 AACCAAGATGTTGTTGCTCACTGTTGACAGAGCTCGGAATCCGCTGTATACACAGC 13191  
QY 1321 ACGCAAGCTCTCACTGCTGTGATCATCACTTACAGACAGATGGGAGCTCGT 1380  
DB 13192 ACGCAAGCTCTCACTGCTGTGATCATCACTTACAGACAGATGGGAGCTCGT 13251  
QY 1381 CTGCGTGGGAATGCTGACGAGAGTTCGCGCGGACCTTGTGAGACATGCTCGAA 1440  
DB 13252 CTGCGTGGGAATGCTGACGAGAGTTCGCGCGGACCTTGTGAGACATGCTCGAA 13311  
QY 1441 GGTACGTGTTTCTCGGCGGCTCACTGAGAAACATGAGGCTGAGAGGCTGCTGT 1500  
DB 13312 GGTACGTGTTTCTCGGCGGCTCACTGAGAAACATGAGGCTGAGAGGCTGCTGT 13371  
QY 1501 TCGCTTCCGCTGCGGAGTAAAGCGGAGCGGAGCGCAACCGCATCCGCTGCTG 1560  
DB 13372 TCGCTTCCGCTGCGGAGTAAAGCGGAGCGGAGCGCAACCGCATCCGCTGCTG 13431

QY 1561 AGCGACATACGCTGCAAGCGCTGTTTCGCGCGCGGCTGAGACGCTGCCATGAGCTC 1620  
DB 13432 AGCGAGCATACGCTGCAAGCGCTGTTTCGCGCGCGGCTGAGACGCTGCCATGAGCTC 13491  
QY 1621 GCCGTGTGTGCGCGCGCAAGCGCTCACGTACGAAGAGCTTTCGCGCGCTGCGCGGCA 1680  
DB 13492 GCCGTGTGTGCGCGCGCAAGCGCTCACGTACGAAGAGCTTTCGCGCGCTGCGCGGCA 13551  
QY 1661 CTTGCGCGCGCGCTGCGCGAGCGAGGCGCGCGCGCAACATTTGTGCGCGGTGTGATG 1740  
DB 13552 CTTGCGCGCGCGCTGCGCGAGCGAGGCGCGCGCGCAACATTTGTGCGCGGTGTGATG 13611  
QY 1741 GAGAAAGCTGCGGAGCAGAGTTTTCGCGGTTCGCGGTGCTGAGTACAGCGCGCTAC 1800  
DB 13612 GAGAAAGCTGCGGAGCAGAGTTTTCGCGGTTCGCGGTGCTGAGTACAGCGCGCTAC 13671  
QY 1801 GTGCCGATCGATCCGACCTTACCGCGGAGCGATTCACCTTCCTCGATCATGTGAG 1860  
DB 13672 GTGCCGATCGATCCGACCTTACCGCGGAGCGATTCACCTTCCTCGATCATGTGAG 13731  
QY 1861 GTAAAGCTGTGCTGACGAGCATGCGCTGATGCGAACTGTCAATGCGCGCGGAGATC 1920  
DB 13732 GTAAAGCTGTGCTGACGAGCATGCGCTGATGCGAACTGTCAATGCGCGCGGAGATC 13791  
QY 1921 CAGCGGCTGCTGCTGAGCGAGCGCGCGCTCGAAGCGAGCGAGCGAGCTCCGATGATG 1980  
DB 13792 CAGCGGCTGCTGCTGAGCGAGCGCGCGCTCGAAGCGAGCGAGCGAGCTCCGATGATG 13851  
QY 1961 CCCATTCAAGACCTTTCGATCTGCGCTGATGCTCATCTACCTTCGCGGATTCACAGGTTG 2040  
DB 13852 CCCATTCAAGACCTTTCGATCTGCGCTGATGCTCATCTACCTTCGCGGATTCACAGGTTG 13911  
QY 2041 CCCAAGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100  
DB 13912 CCCAAGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13971  
QY 2101 CGCTTCGAATATGAGCGCGCGAGACAGAGGCTGCGCTCTCTCGCTGAGTTTCATCTC 2160  
DB 13972 CGCTTCGAATATGAGCGCGCGAGACAGAGGCTGCGCTCTCTCGCTGAGTTTCATCTC 14031  
QY 2161 TCGGCTCATGATGTGTTGGGATCTGCGCGCGCGCGGCGGATCGATCTGTGTCGAGCGG 2220  
DB 14032 TCGGCTCATGATGTGTTGGGATCTGCGCGCGCGCGGCGGATCGATCTGTGTCGAGCGG 14091  
QY 2221 TCCAAGCTGCGGATCCGCGCGATGCGGAGGATGATGATGATGATGATGATGATGATGATG 2280  
DB 14092 TCCAAGCTGCGGATCCGCGCGATGCGGAGGATGATGATGATGATGATGATGATGATG 14151  
QY 2281 TGAAGCTGAGTCCGCGCGCTGATGCGGATGCTGTCGAGCATTTTGAAGGTCGCCCGAT 2340  
DB 14152 TGAAGCTGAGTCCGCGCGCTGATGCGGATGCTGTCGAGCATTTTGAAGGTCGCCCGAT 14211  
QY 2341 TCGCTCGTAAAGTCTCTGCGGCTTTCGCTGTCGAGCGCGGATGATCCCGTGGGCGCTG 2400  
DB 14212 TCGCTCGTAAAGTCTCTGCGGCTTTCGCTGTCGAGCGCGGATGATCCCGTGGGCGCTG 14271  
QY 2401 CTTGCGCGAGCTTCCAGGCGCATGAGCGCGCGCTGTGCTGATGATGAGCTTGGGCGGCGAC 2460  
DB 14272 CTTGCGCGAGCTTCCAGGCGCATGAGCGCGCGCTGTGCTGATGATGAGCTTGGGCGGCGAC 14331  
QY 2461 GAAAGCTGATGTGCTCATCGGATACCGCTGAGAAAGCTGATGATGATGATGATGATGATG 2520  
DB 14332 GAAAGCTGATGTGCTCATCGGATACCGCTGAGAAAGCTGATGATGATGATGATGATGATG 14391  
QY 2521 ATCCCTTACGCGCGCTGCTGCGCAACGAGCGTTTCACGTCGTCGATGAGCGCTCGAA 2580  
DB 14392 ATCCCTTACGCGCGCTGCTGCGCAACGAGCGTTTCACGTCGTCGATGAGCGCTCGAA 14451  
QY 2581 CCGCGCGCGCTGCTGCGGCGAATCTTACATTTGCGGCGTGGGCTGGCCTGCGACCTGGG 2640  
DB 14452 CCGCGCGCGCTGCTGCGGCGAATCTTACATTTGCGGCGTGGGCTGGCCTGCGACCTGGG 14511

QY 2641 TACTGGCGCGATGAAGAGAGACGCGCAAGAGCTTCTCTGTCACCCCGAGACCGGGAG 2700  
DB 14512 TACTGGCGCGATGAAGAGAGACGCGCAAGAGCTTCTCTGTCACCCCGAGACCGGGAG 14571  
QY 2701 GCGCTTTCAGAGCGGAGATCTGGGCGGCTACCTGCGCGGATGAGAAATAGAGTTTCATG 2760  
DB 14572 GCGCTTTCAGAGCGGAGATCTGGGCGGCTACCTGCGCGGATGAGAAATAGAGTTTCATG 14631  
QY 2761 GGGCGTGAAGACCAACCAATCAAGCTTTCGCGGATACCGCTTGAAGCTCGGGAAATCGAG 2820  
DB 14632 GGGCGTGAAGACCAACCAATCAAGCTTTCGCGGATACCGCTTGAAGCTCGGGAAATCGAG 14691  
QY 2821 GAAAGCTCAAGTGCATCCGACGTAACGCGAGCGGATGATGATGATGATGATGATGATGATG 2880  
DB 14692 GAAAGCTCAAGTGCATCCGACGTAACGCGAGCGGATGATGATGATGATGATGATGATG 14751  
QY 2881 GCGGCGAACAAGCTTCTTTCGCTTATGCTTATGCTTCCCGAGGCGACACCGAGACCGCTCC 2940  
DB 14752 GCGGCGAACAAGCTTCTTTCGCTTATGCTTATGCTTCCCGAGGCGACACCGAGACCGCTCC 14811  
QY 2941 GAGCGAGACGCGAGCTCAAGACCGAGCGGATTCGACGCGAGACACGCGCGCGAAGCG 3000  
DB 14812 GAGCGAGACGCGAGCTCAAGACCGAGCGGATTCGACGCGAGACACGCGCGCGAAGCG 14871  
QY 3001 GACGCTTGAACGACGCGGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 3060  
DB 14872 GACGCTTGAACGACGCGGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 14931  
QY 3061 GACCTGAGCGGAAAGCCGCTGCTGATTTGACCGGCGAGATTCGCGGAGAGCGGGCTG 3120  
DB 14932 GACCTGAGCGGAAAGCCGCTGCTGATTTGACCGGCGAGATTCGCGGAGAGCGGGCTG 14991  
QY 3121 GACGCTTGAACGCGGCGGCTGAGGCTGCGAAGTTCTCTTGAAGCGCGGATTCGCTTGT 3180  
DB 14992 GACGCTTGAACGCGGCGGCTGAGGCTGCGAAGTTCTCTTGAAGCGCGGATTCGCTTGT 15051  
QY 3181 GAGTTTGTGCTGATCTGAGCTGCTTGAAGACGCTGAGCGCGGACCGGAGCGGACCTTCC 3240  
DB 15052 GAGTTTGTGCTGATCTGAGCTGCTTGAAGACGCTGAGCGCGGAGCGGAGCGGACCTTCC 15111  
QY 3241 AAATTCGCTTATCCATCGCGCGGCGAGCATCCCGGTGCAAACTTACGCTATGTCAAA 3300  
DB 15112 AAATTCGCTTATCCATCGCGCGGCGAGCATCCCGGTGCAAACTTACGCTATGTCAAA 15171  
QY 3301 TCCGCGCGCATGAGGCGGTGAGACGAGGCTTCTTATTTACACCGCTTGAGACCGT 3360  
DB 15172 TCCGCGCGCATGAGGCGGTGAGACGAGGCTTCTTATTTACACCGCTTGAGACCGT 15231  
QY 3361 TTGCTGAAGCTCTCGATCAAGGATCGAGCGCGAGCGCATGTTCCGCAAAACTTCGAC 3420  
DB 15232 TTGCTGAAGCTCTCGATCAAGGATCGAGCGCGAGCGCATGTTCCGCAAAACTTCGAC 15291  
QY 3421 GTGTTGATGAAGCGGCGTTCACCTCTGTTTCGTTGCGAGATGACCGCATGAGTTCG 3480  
DB 15292 GTGTTGATGAAGCGGCGTTCACCTCTGTTTCGTTGCGAGATGACCGCATGAGTTCG 15351  
QY 3481 CTGATGATGATGCTGCGGAGAAATTTGCGTGTGAGGCGGATATATGCGGAGCTC 3540  
DB 15352 CTGATGATGATGCTGCGGAGAAATTTGCGTGTGAGGCGGATATATGCGGAGCTC 15411  
QY 3541 CTGATGAGCAGGCGCTTCTGCAACATCGGCGTCTGTCGCTGGGCGCAATTCATTTT 3600  
DB 15412 CTGATGAGCAGGCGCTTCTGCAACATCGGCGTCTGTCGCTGGGCGCAATTCATTTT 15471  
QY 3601 GAAAGGTTGCGCGGCTTCTGACCTGCGACATTTGAGAGCTTTTACGTCACGCGATGCTG 3660  
DB 15472 GAAAGGTTGCGCGGCTTCTGACCTGCGACATTTGAGAGCTTTTACGTCACGCGATGCTG 15531  
QY 3661 GCGCGGCGGATGAGACCGCGCAAGTTTCAGGCTGTCAGCTCGGTCAGGATTCCTCACCG 3720  
DB 15532 GCGCGGCGGATGAGACCGCGCAAGTTTCAGGCTGTCAGCTCGGTCAGGATTCCTCACCG 15591  
QY 3721 AGCGCGCGCACGAGCGCGCGCGCTTCCGCGCGCGAGCAGACTTTCGCGATATGCTT 3780





Db 13072 GCGCGGCTCTGGGGATCCAGAGCGCATGTTCCCGTGTCTCAGCAGCGCGCTT 13131  
Qy 1261 AACGAGCAAGTCGTTGGTGTCACTCGTTGAGAGGCTCGGAACTCCGGTGTACACAGC 1320  
Db 13132 AACGAGCAAGTCGTTGGTGTCACTCGTTGAGAGGCTCGGAACTCCGGTGTACACAGC 13191  
Qy 1321 ACGGAGACTCCTCAGAGCTGCTGTGATCATCAGCTCTACAGACGATGGGAGCTCGTC 1380  
Db 13192 ACGGAGACTCCTCAGAGCTGCTGTGATCATCAGCTCTACAGACGATGGGAGCTCGTC 13251  
Qy 1381 CTCGCTGGGACATGTCGACGAGAGTGTCCCGCCGACTTCTGAGACGATGCTCGAA 1440  
Db 13252 CTCGCTGGGACATGTCGACGAGTGTCCCGCCGACTTCTGAGACGATGCTCGAA 13311  
Qy 1441 GCGTACGTCGTTTCTCCGGGCTCTAGAGAACATGGGGTGAACAGGTGGCTGT 1500  
Db 13312 GCGTACGTCGTTTCTCCGGGCTCTAGAGAACATGGGGTGAACAGGTGGCTGT 13371  
Qy 1501 TCGCTCCGCTGCGCAGCTAGAGAGCGGGCGAGAGCAACGCGACTGCTG 1560  
Db 13372 TCGCTCCGCTGCGCAGCTAGAGAGCGGGCGAGAGCAACGCGACTGCTG 13431  
Qy 1561 AGCGAGCATACGTCGACGCGCTGTTCCGGCGGGTTCAGACAGCTGCCATGACGCTC 1620  
Db 13432 AGCGAGCATACGTCGACGCGCTGTTCCGGCGGGTTCAGACAGCTGCCATGACGCTC 13491  
Qy 1621 GCGCTGTGTGGCGCGGCAACGCTCACTACAGAGCTTTCCGCCCTTCGGGGGA 1680  
Db 13492 GCGCTGTGTGGCGCGGCAACGCTCACTACAGAGCTTTCCGCCCTTCGGGGGA 13551  
Qy 1681 CTTGGCGCGGCTGCGGAGCAGAGGGGAGCGCCGGAACATTTGTCGGGTGGTAG 1740  
Db 13552 CTTGGCGCGGCTGCGGAGCAGAGGGGAGCGCCGGAACATTTGTCGGGTGGTAG 13611  
Qy 1741 GAGAAAGCTGGAGCAGGTTGTGCGGTTCTCGGCTGCTGAGTCAGGCGCGCTAC 1800  
Db 13612 GAGAAAGCTGGAGCAGGTTGTGCGGTTCTCGGCTGCTGAGTCAGGCGCGCTAC 13671  
Qy 1801 GTGCGCATGATGCCGACCTACCGCGGAGAGGCTACCTACCTCTCTGATCATGTGTAG 1860  
Db 13672 GTGCGCATGATGCCGACCTACCGCGGAGAGGCTACCTACCTCTCTGATCATGTGTAG 13731  
Qy 1861 GTAAAGCTGTCGTGACGACGATGCTGTGATGGCAACTGTCACTGCGCGGGATC 1920  
Db 13732 GTAAAGCTGTCGTGACGACGATGCTGTGATGGCAACTGTCACTGCGCGGGATC 13791  
Qy 1921 CAGCGGCTGCTGTGAGCGAGGCGCGGTTCGAAAGCGACGCGACCTTCGATGATG 1980  
Db 13792 CAGCGGCTGCTGTGAGCGAGGCGCGGTTCGAAAGCGACGCGACCTTCGATGATG 13851  
Qy 1981 CCCATTGAGACACCTTCGGAATCTCGCGTATGTCACTCACTCGGGATTCACAGGGTTC 2040  
Db 13852 CCCATTGAGACACCTTCGGAATCTCGCGTATGTCACTCACTCGGGATTCACAGGGTTC 13911  
Qy 2041 CCCAAGGGGGTGTATGATCATCGGGGTGCGCTCAACACCATCTGACATCAACAG 2100  
Db 13912 CCCAAGGGGGTGTATGATCATCGGGGTGCGCTCAACACCATCTGACATCAACAG 13971  
Qy 2101 CGCTTCGAATAGGCGCCGAGACAGGGTGTGCGCTCTCTCGCTAGGCTTCGATCTC 2160  
Db 13972 CGCTTCGAATAGGCGCCGAGACAGGGTGTGCGCTCTCTCGCTAGGCTTCGATCTC 14031  
Qy 2161 TCGGCTATGATGTGTTGGGATCTTGCGCGCGGGGCTGATGTGTGCGGACCG 2220  
Db 14032 TCGGCTATGATGTGTTGGGATCTTGCGCGCGGGGCTGATGTGTGCGGACCG 14091  
Qy 2221 TCCAAGTCGCGGATCCGGCGATTTGGGAGATTGATCGAACGAGAGAGGTGACGGTG 2280  
Db 14092 TCCAAGTCGCGGATCCGGCGATTTGGGAGATTGATCGAACGAGAGAGGTGACGGTG 14151  
Qy 2281 TGGAACTGGGTGCGCGCTGATGCTGTCGAGCAATTTGAGGGTCGCCCGAT 2340

Db 14152 TGGAACTGGGTGCGCGCTGATGCTGTCGAGCAATTTGAGGGTCGCCCGAT 14211  
Qy 2341 TCGCTCGTATGATCTCTGCGGCTTTCGCTGTGAGCGGCGCATGAGATCCGGTGGGCTG 2400  
Db 14212 TCGCTCGTATGATCTCTGCGGCTTTCGCTGTGAGCGGCGCATGAGATCCGGTGGGCTG 14271  
Qy 2401 CTTGGCAGCTCCAGGCGCATCAGGCGCGGCTGTGTGATCATCAGCTTGGCGGGGCGCAC 2460  
Db 14272 CTTGGCAGCTCCAGGCGCATCAGGCGCGGCTGTGTGATCATCAGCTTGGCGGGGCGCAC 14331  
Qy 2461 GAAGCTGATCTGTGCTCATTCGGGTACCCCGTAGAGAACGTTCAGACTATCTGGCGAGC 2520  
Db 14332 GAAGCTGATCTGTGCTCATTCGGGTACCCCGTAGAGAACGTTCAGACTATCTGGCGAGC 14391  
Qy 2521 ATCCCTACCGGCGCTTCGGCTGCGCAACGAGCTTCCACGCTGCTGATGAGCGCTCGAA 2580  
Db 14392 ATCCCTACCGGCGCTTCGGCTGCGCAACGAGCTTCCACGCTGCTGATGAGCGCTCGAA 14451  
Qy 2581 CCGCGCGGCTTGGGTTCCGGGCGAACTTCAATTGGCGGGTTCGGCTGGCACTGGGC 2640  
Db 14452 CCGCGCGGCTTGGGTTCCGGGCGAACTTCAATTGGCGGGTTCGGCTGGCACTGGGC 14511  
Qy 2641 TACTGGCGCATTAAGAGAGACGCGCAAGCTTCTCTGTGACCCCGAGACCGGGAG 2700  
Db 14512 TACTGGCGCATTAAGAGAGACGCGCAAGCTTCTCTGTGACCCCGAGACCGGGAG 14571  
Qy 2701 GCGCTTACAAAGCGCGGATCTGGGCGGCTACCTGCGGATGGAACATGAGTTATG 2760  
Db 14572 GCGCTTACAAAGCGCGGATCTGGGCGGCTACCTGCGGATGGAACATGAGTTATG 14631  
Qy 2761 GGGCGTAGAGCAACCAATCAAGCTTCCGCGATACCGCGTTGAGCTCGGGAATCGAG 2820  
Db 14632 GGGCGTAGAGCAACCAATCAAGCTTCCGCGATACCGCGTTGAGCTCGGGAATCGAG 14691  
Qy 2821 GAAGCTCAAGTTCGATTCGGAACGTACCGGATACCGCGGATTTGCTCGGAGACAC 2880  
Db 14692 GAAGCTCAAGTTCGATTCGGAACGTACCGGATACCGCGGATTTGCTCGGAGACAC 14751  
Qy 2881 GCGGCGAAAGGCTCTTACCTATGCTATGCTGCGGAGGCGACAGGAGCGGGCTGC 2940  
Db 14752 GCGGCGAAAGGCTCTTACCTATGCTATGCTGCGGAGGCGACAGGAGCGGGCTGC 14811  
Qy 2941 GAGCAGACGCGAGCTTCAAGACCGAGCGGATTCAGCGGAGACACGCGCGGAAGCG 3000  
Db 14812 GAGCAGACGCGAGCTTCAAGACCGAGCGGATTCAGCGGAGACACGCGCGGAAGCG 14871  
Qy 3001 GACGCTTGAAGCGCGGAGAGGGTCAATTCAAGCTTCGTCGACACGACTTCGGAGG 3060  
Db 14872 GACGCTTGAAGCGCGGAGAGGGTCAATTCAAGCTTCGTCGACACGACTTCGGAGG 14931  
Qy 3061 GACCTGAGCGGAAAGCGGCTGTGATCTGACCGGGGAGGATTCGGGGAGGGGGCTG 3120  
Db 14932 GACCTGAGCGGAAAGCGGCTGTGATCTGACCGGGGAGGATTCGGGGAGGGGGCTG 14991  
Qy 3121 GACGCTACGCGCGCTGCGCGTACGCTCGGAAGCTTCTTGAAGGCCCGATTCCGTTGTT 3180  
Db 14992 GACGCTACGCGCGCTGCGCGTACGCTCGGAAGCTTCTTGAAGGCCCGATTCCGTTGTT 15051  
Qy 3181 GAGTTGGTGTATCTGAGCTGTTAGCAGCGTGAAGCCGAGCGGCGGACCTTCC 3240  
Db 15052 GAGTTGGTGTATCTGAGCTGTTAGCAGCGTGAAGCCGAGCGGCGGACCTTCC 15111  
Qy 3241 AAATTCGTTATTCATGCGGGGAGAGACGTAACCGGTCGAAACCTACGCTATGTCAA 3300  
Db 15112 AAATTCGTTATTCATGCGGGGAGAGACGTAACCGGTCGAAACCTACGCTATGTCAA 15171  
Qy 3301 TCGGCGCATTCGAGGCGTGTGAGAGGGCTTCAATTAATCAACCGCTTGAACACCGT 3360  
Db 15172 TCGGCGCATTCGAGGCGTGTGAGAGGGCTTCAATTAATCAACCGCTTGAACACCGT 15231  
Qy 3361 TTGCTGAAGCTTCCGATTCAGCGGATTCAGCGCGGAGCGGACGCTTGGCAAACTTGAC 3420  
Db 15232 TTGCTGAAGCTTCCGATTCAGCGGATTCAGCGCGGAGCGGACGCTTGGCAAACTTGAC 15291



QY 3421 GTGTTGATGAAGCGGCTTCAACTCTGTTGTCGGGAGATGACCCATCGAGTCG 3480  
DB 15292 GTTTCGATGAAGCGGCTTCAACTCTGTTGTCGGGAGATGACCCATCGAGTCG 15351  
QY 3481 CTGATGATGATCGTCGTCGAGAAATTTGCTGTCGGAGGCGGGAATATGAGGAGCTC 3540  
DB 15352 CTGATGATGATCGTCGTCGAGAAATTTGCTGTCGGAGGCGGGAATATGAGGAGCTC 15411  
QY 3541 CTGATGATGATCGTCGTCGAGAAATTTGCTGTCGGAGGCGGGAATATGAGGAGCTC 3600  
DB 15412 CTGATGATGATCGTCGTCGAGAAATTTGCTGTCGGAGGCGGGAATATGAGGAGCTC 15471  
QY 3601 GAAACAGTTTCGGCGGCTTCTCGACTGCACTTCGACGTTTACCTGCAACGCGATGTC 3660  
DB 15472 GAAACAGTTTCGGCGGCTTCTCGACTGCACTTCGACGTTTACCTGCAACGCGATGTC 15531  
QY 3661 GGGGGGGGGGTGAGACCGGGGGAGTTTCAAGTCTGTACGCTCGGTGAGATTCTTCACCG 3720  
DB 15532 GGGGGGGGGGTGAGACCGGGGGAGTTTCAAGTCTGTACGCTCGGTGAGATTCTTCACCG 15591  
QY 3721 AGCGCGCCACGACGCGCGCGCGCCCTCCCGCGCGAGACGACTTCGCGATATGCTT 3780  
DB 15592 AGCGCGCCACGACGCGCGCGCGCCCTCCCGCGCGAGACGACTTCGCGATATGCTT 15651  
QY 3781 CGCGACTTCTTGAGACCAAACTAACCCGAGTACATGTCCTTCAAGTCTTCTGAGACTC 3840  
DB 15652 CGCGACTTCTTGAGACCAAACTAACCCGAGTACATGTCCTTCAAGTCTTCTGAGACTC 15711  
QY 3841 GATGCGTTCGCGTGAAGTCAAGGAGGATGATGATGATGATGATGATGATGATGATGATG 3900  
DB 15712 GATGCGTTCGCGTGAAGTCAAGGAGGATGATGATGATGATGATGATGATGATGATGATG 15771  
QY 3901 GATGCGTTCGCGTGAAGTCAAGGAGGATGATGATGATGATGATGATGATGATGATGATG 3960  
DB 15772 GATGCGTTCGCGTGAAGTCAAGGAGGATGATGATGATGATGATGATGATGATGATGATG 15831  
QY 3961 CTGCTGCGCGCTGCTACGAGAGTGTCTCGGCTGAGAGTGTCTCGGCTTCCAGAGACTTC 4020  
DB 15832 CTGCTGCGCGCTGCTACGAGAGTGTCTCGGCTGAGAGTGTCTCGGCTTCCAGAGACTTC 15891  
QY 4021 GTGCGATCTTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4080  
DB 15892 GTGCGATCTTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 15951  
QY 4081 CTGAGATGAGGAGATGCGCATCAACGAGTTTTCAGTACCCGAACTCGGCTGCTGCGG 4140  
DB 15952 CTGAGATGAGGAGATGCGCATCAACGAGTTTTCAGTACCCGAACTCGGCTGCTGCGG 16011  
QY 4141 TCCGCTTTCGCGAGACATCTGAGAGATCTAGATCAACGCGCCGAACTGAGAGCCGAGTG 4200  
DB 16012 TCCGCTTTCGCGAGACATCTGAGAGATCTAGATCAACGCGCCGAACTGAGAGCCGAGTG 16071  
QY 4201 GAGGTTCCGCGCAAGGCGAGAGACGTAGCTAA 4233  
DB 16072 GAGGTTCCGCGCAAGGCGAGAGACGTAGCTAA 16104

RESULT 6  
US-09-568-472-1  
; Sequence 1, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,472  
; CURRENT FILING DATE: 2000-05-10

;; PRIOR APPLICATION NUMBER: 09/335,409  
;; PRIOR FILING DATE: 1999-06-17  
;; NUMBER OF SEQ ID NOS: 30  
;; SOFTWARE: Patentin Ver. 2.0  
;; SEQ ID NO 1  
;; LENGTH: 68750  
;; TYPE: DNA  
;; ORGANISM: Sorangium cellulosum  
US-09-568-472-1  
Query Match 100.0%; Score 4233; DB 4; Length 68750;  
Beef Local Similarity 100.0%; Pred. No. 0;  
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAGATCAATCAGCTTCTGAACGAGCTGACGACGAGGATATCACTGCGGCGCAT 60  
DB 11872 ATGAGATCAATCAGCTTCTGAACGAGCTGACGACGAGGATATCACTGCGGCGCAT 11931  
QY 61 GGGGAGCGGCTTCAGATACAGGCGCCGCAAGAACGCGTGAACCGGACCTGCTGCGA 120  
DB 11932 GGGGAGCGGCTTCAGATACAGGCGCCGCAAGAACGCGTGAACCGGACCTGCTGCGA 11991  
QY 121 ATCTCGAGCAAAAGACGATCTGACGATGCTCTGACAGACTCCCGCAGAAATCC 180  
DB 11992 ATCTCGAGCAAAAGACGATCTGACGATGCTCTGACAGACTCCCGCAGAAATCC 12051  
QY 181 ATGTCGCCGCCGCGAGCGGAGCGGCTGCTGCTCTGACAGACATCCAGAAATCC 240  
DB 12052 ATGTCGCCGCCGCGAGCGGAGCGGCTGCTGCTCTGACAGACATCCAGAAATCC 12111  
QY 241 TACTGCGTGGCGGAGACAGAGCGGTTTACGGTCCGACGCGGATCCAGCGCTATCGGAA 300  
DB 12112 TACTGCGTGGCGGAGACAGAGCGGTTTACGGTCCGACGCGGATCCAGCGCTATCGGAA 12171  
QY 301 TACGACTGAGGATCTGACGCTGCGAGGCTGAGCGCGCTTTCGAGAAATGCTGCGG 360  
DB 12172 TACGACTGAGGATCTGACGCTGCGAGGCTGAGCGCGCTTTCGAGAAATGCTGCGG 12231  
QY 361 CGGACGACATGCTTTCGCGCCCAACGCTGCGCCGACATGATGACGATGAGGCTTAA 420  
DB 12232 CGGACGACATGCTTTCGCGCCCAACGCTGCGCCGACATGATGACGATGAGGCTTAA 12291  
QY 421 GTGAGCGCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
DB 12292 GTGAGCGCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12351  
QY 481 AGGCTGATGCTGTCGAGATGCGATGTCGACCGCATCTATGACACCGAGCGCTTCG 540  
DB 12352 AGGCTGATGCTGTCGAGATGCGATGTCGACCGCATCTATGACACCGAGCGCTTCG 12411  
QY 541 CTCTATCAAGTGTGCGGCTTTCGCTGAGACGAGCGGCAAAACCGCTGCTGCTGATGTC 600  
DB 12412 CTCTATCAAGTGTGCGGCTTTCGCTGAGACGAGCGGCAAAACCGCTGCTGCTGATGTC 12471  
QY 601 GATCTCATTAACGTTGACCTAGGAGCGCTGTCATCATCTTCAAGGCTGCTGAGCTTC 660  
DB 12472 GATCTCATTAACGTTGACCTAGGAGCGCTGTCATCATCTTCAAGGCTGCTGAGCTTC 12531  
QY 661 TACGAGATCCGAGACCTTCTCTCTGTCGAGCTCTGATCCGAGTTATGATCTC 720  
DB 12532 TACGAGATCCGAGACCTTCTCTCTGTCGAGCTCTGATCCGAGTTATGATCTC 12591  
QY 721 GCGCTGAGTCTCGCAAGAGTCTGAGCGCATCAACGATGATGATGATGATGATGATGATGATG 780  
DB 12592 GCGCTGAGTCTCGCAAGAGTCTGAGCGCATCAACGATGATGATGATGATGATGATGATGATG 12651  
QY 781 CGCATGCGAGACTCCACCTCCGCGAGGCTTTCGATGAGGCGGATCTCACTACCTG 840  
DB 12652 CGCATGCGAGACTCCACCTCCGCGAGGCTTTCGATGAGGCGGATCTCACTACCTG 12711  
QY 841 AAGGAGATCCGCTTCGCGACACGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
DB 12712 AAGGAGATCCGCTTCGCGACACGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12771

QY 901 AAGCGGCTGTTCGGGAGCGGGGCTGAACCCGACGGGCTCATCTGCTGATTTTCC 960  
Db 12772 AAGCGGCTGTTCGGGAGCGGGGCTGAACCCGACGGGCTCATCTGCTGATTTTCC 12831  
QY 961 GAGGTGATCGGGGCTGTGAGCGCGAGCCCGGTTTACGCTCAATAAACGCTTCAAC 1020  
Db 12832 GAGGTGATCGGGGCTGTGAGCGCGAGCCCGGTTTACGCTCAATAAACGCTTCAAC 12891  
QY 1021 CGGCTCCCGGTCCATCGGGGCTGAACGATATCAACGGGGGCTTACGCTCAATGCTCTC 1080  
Db 12892 CGGCTCCCGGTCCATCGGGGCTGAACGATATCAACGGGGGCTTACGCTCAATGCTCTC 12951  
QY 1081 CTGACATCGACACCACTCGCGACAGAGCTTCAAGACGGCGCTAAAGCTTCAAGAG 1140  
Db 12952 CTGACATCGACACCACTCGCGACAGAGCTTCAAGACGGCGCTAAAGCTTCAAGAG 13011  
QY 1141 CAGCTGTGGGAAGCATGATATCATCTGCAAGCTTACGCTTCAAGAGAGCC 1200  
Db 13012 CAGCTGTGGGAAGCATGATATCATCTGCAAGCTTACGCTTCAAGAGAGCC 13071  
QY 1201 GCCCGGCTCTGGGGATCCACAGAGGCGCATTTTCCCGTGTGCTCAAGAGCGGCTT 1260  
Db 13072 GCCCGGCTCTGGGGATCCACAGAGGCGCATTTTCCCGTGTGCTCAAGAGCGGCTT 13131  
QY 1261 AACCAAGAGTCTGTGTGTCACTCTGTGACAGAGCTCGAACTCCGCTGTACACAGC 1320  
Db 13132 AACCAAGAGTCTGTGTGTCACTCTGTGACAGAGCTCGAACTCCGCTGTACACAGC 13191  
QY 1321 ACGCAGACTCTCAGCTGCTGTGATCATCATCTTACAGACGATGGGAGCCTCTCTC 1380  
Db 13192 ACGCAGACTCTCAGCTGCTGTGATCATCATCTTACAGACGATGGGAGCCTCTCTC 13251  
QY 1381 CTCGGTGGGACATGTGACAGAGAGTCTCCCGCCGACCTTTCGAGACGATGCTCGAA 1440  
Db 13252 CTCGGTGGGACATGTGACAGAGAGTCTCCCGCCGACCTTTCGAGACGATGCTCGAA 13311  
QY 1441 GCGTACGTCTGTCTTCTCCGGGCGCTCATCTGAGAACATGGGAGTGAACAGGTGCTGT 1500  
Db 13312 GCGTACGTCTGTCTTCTCCGGGCGCTCATCTGAGAACATGGGAGTGAACAGGTGCTGT 13371  
QY 1501 TCGCTTCCGCTGCCAGCTTGAAGCGCGGCGAGCGCAACGCGACCAACGCGCTGCTG 1560  
Db 13372 TCGCTTCCGCTGCCAGCTTGAAGCGCGGCGAGCGCAACGCGACCAACGCGCTGCTG 13431  
QY 1561 AGCGAGCATAGCTGCGCAAGGCTGTTCGGGCGCGGCTGACAGCTCCATGAGCTC 1620  
Db 13432 AGCGAGCATAGCTGCGCAAGGCTGTTCGGGCGCGGCTGACAGCTCCATGAGCTC 13491  
QY 1621 GCCGTGTGTGCGGCGCGCAAGCGCTCATCTGACAAAGCTTTCGCGCTTCCGGGCGA 1680  
Db 13492 GCCGTGTGTGCGGCGCGCAAGCGCTCATCTGACAAAGCTTTCGCGCTTCCGGGCGA 13551  
QY 1681 CTTGCGCGGCGGCTGCGGAGAGGAGGCAAGCCCGAAACATTTGTCGCGTGTGATG 1740  
Db 13552 CTTGCGCGGCGGCTGCGGAGAGGAGGCAAGCCCGAAACATTTGTCGCGTGTGATG 13611  
QY 1741 GAGAAAGGCTGGAGCAGGTTGTCCGGCTTCTCGGCTGTGAGTCAAGGCGGCTTAC 1800  
Db 13612 GAGAAAGGCTGGAGCAGGTTGTCCGGCTTCTCGGCTGTGAGTCAAGGCGGCTTAC 13671  
QY 1801 GTGCGATCGATGCGACCTTACCGGCGAGGCTATCCACTCTCTGATCATGTTAG 1860  
Db 13672 GTGCGATCGATGCGACCTTACCGGCGAGGCTATCCACTCTCTGATCATGTTAG 13731  
QY 1861 GTAAAGCTGTGTGACGACGATGCTGTGATGCAATCTGTCAATGCGCGCGGAGATC 1920  
Db 13732 GTAAAGCTGTGTGACGACGATGCTGTGATGCAATCTGTCAATGCGCGCGGAGATC 13791  
QY 1921 CAGCGGCTGTGTGAGGAGGCGCGCTGCAAGGCGACGCGACCACTTCCGATGATG 1980  
Db 13792 CAGCGGCTGTGTGAGGAGGCGCGCTGCAAGGCGACGCGACCACTTCCGATGATG 13851

QY 1981 CCATTCAGACACTTCGGATCTGCGTATGTCATTCACCTCGGATCCAGAGGTTG 2040  
Db 13852 CCATTCAGACACTTCGGATCTGCGTATGTCATTCACCTCGGATCCAGAGGTTG 13911  
QY 2041 CCAGAGGGGTGATGATTCATTCAGGAGTCCGTCAACACATCTTGGATCATCAAGAG 2100  
Db 13912 CCAGAGGGGTGATGATTCATTCAGGAGTCCGTCAACACATCTTGGATCATCAAGAG 13971  
QY 2101 CGTTTCAATATAGGGCCCGAGACAGGGTGTGGCGCTCTCTGCTAGGCTTCAATCTC 2160  
Db 13972 CGTTTCAATATAGGGCCCGAGACAGGGTGTGGCGCTCTCTGCTAGGCTTCAATCTC 14031  
QY 2161 TCGTCTATGATGTGTTTCGGATCTGTGCGGCGGAGCTGATGATGCTGTCGAGCGC 2220  
Db 14032 TCGTCTATGATGTGTTTCGGATCTGTGCGGCGGAGCTGATGATGCTGTCGAGCGC 14091  
QY 2221 TCAGAGTGGCGGATTCGGGCGCATTTGGGCAAGTTGATTCAGACGAGAGAGTGAAGG 2280  
Db 14092 TCAGAGTGGCGGATTCGGGCGCATTTGGGCAAGTTGATTCAGACGAGAGAGTGAAGG 14151  
QY 2281 TGAAGCTCGGTGCGGGGCTGATGCGGATGCTGCTGAGCATTTTGAAGGCTGCGCCGAT 2340  
Db 14152 TGAAGCTCGGTGCGGGGCTGATGCGGATGCTGCTGAGCATTTTGAAGGCTGCGCCGAT 14211  
QY 2341 TCGCTCGTATAGTCTTGTGCGGCTTTCGCTGTGAGCGGCGACTGATTCGGGTGGCTG 2400  
Db 14212 TCGCTCGTATAGTCTTGTGCGGCTTTCGCTGTGAGCGGCGACTGATTCGGGTGGCTG 14271  
QY 2401 CTTGCGAGCTTCAGGCGCATGAGCCCGGCGTGTGCGTATGATGATCTGGGCGGCGCAC 2460  
Db 14272 CTTGCGAGCTTCAGGCGCATGAGCCCGGCGTGTGCGTATGATGATCTGGGCGGCGCAC 14331  
QY 2461 GAAAGCTGATCTGTGCTCATGAGGATCCCGTGAAGGAAAGCTGACCTATCGTGGGCGAGC 2520  
Db 14332 GAAAGCTGATCTGTGCTCATGAGGATCCCGTGAAGGAAAGCTGACCTATCGTGGGCGAGC 14391  
QY 2521 ATCCCTTACCGGCGCTGCTGCGGCAACGACGATTCACGCTGATGAGGCGCTCGAA 2580  
Db 14392 ATCCCTTACCGGCGCTGCTGCGGCAACGACGATTCACGCTGATGAGGCGCTCGAA 14451  
QY 2581 CCGGCGCGGCTGTGGGTTCCGGGCAACTTACATTTGGCGGGGTGCGGCTGCACTGGGC 2640  
Db 14452 CCGGCGCGGCTGTGGGTTCCGGGCAACTTACATTTGGCGGGGTGCGGCTGCACTGGGC 14511  
QY 2641 TACTGGGCGATGGAAGGAAGAAGCGGCAAGGCTTCCGTGAGCAACCCGAGACCGGGAG 2700  
Db 14512 TACTGGGCGATGGAAGGAAGAAGCGGCAAGGCTTCCGTGAGCAACCCGAGACCGGGAG 14571  
QY 2701 CGCTCTTACAGACCGGCGATCTGGGCGCTTACCTGCGGATGGAACATCGAGTTCAATG 2760  
Db 14572 CGCTCTTACAGACCGGCGATCTGGGCGCTTACCTGCGGATGGAACATCGAGTTCAATG 14631  
QY 2761 GGGGCTGAGGCAACCAATCAAGCTTTCGCGATACCGCGTTGAGCTTGGGGAATCGAG 2820  
Db 14632 GGGGCTGAGGCAACCAATCAAGCTTTCGCGATACCGCGTTGAGCTTGGGGAATCGAG 14691  
QY 2821 GAAACGCTCAAGTGTGATTCGGAAGTACGCAACCGGCTGATTTGGCCCTGGGGAACAC 2880  
Db 14692 GAAACGCTCAAGTGTGATTCGGAAGTACGCAACCGGCTGATTTGGCCCTGGGGAACAC 14751  
QY 2881 GCGGCGAAAGGCTCTTCTTATGCTTATGTGCTCCGAGAGGCAACAGAGAGCGGCTCC 2940  
Db 14752 GCGGCGAAAGGCTCTTCTTATGCTTATGTGCTCCGAGAGGCAACAGAGAGCGGCTCC 14811  
QY 2941 GAGCAGACGCGAGCTTCAGACCGAGCGATTCAGCGGAGAGCAACGCGCGGAACG 3000  
Db 14812 GAGCAGACGCGAGCTTCAGACCGAGCGATTCAGCGGAGAGCAACGCGCGGAACG 14871  
QY 3001 GACGCTTGAAGCGAGCGGCGAGGCTGCAAGTTCAGCTGCTGCAACGAGCTCCGAGG 3060  
Db 14872 GACGCTTGAAGCGAGCGGCGAGGCTGCAAGTTCAGCTGCTGCAACGAGCTCCGAGG 14931  
QY 3061 GACCTGAGCGGAAAGCCGCTGTGATCTGACCGGAGAGATTCGCGGAGGCGGGCTG 3120

```
DB 14932 GACCTGATGAAAGCCGCTCGTCAATCTGACCGGACAGATCCCGGAGCGGCGCTG 14991
| 14992 GACGCTACGCGCGCTGCGCGCTGAGCGCTCCGAAAGTTCTTGAAGCCCGCAATTCGTTGTT 3180
| 14992 GACGCTACGCGCGCTGCGCGCTGAGCGCTCCGAAAGTTCTTGAAGCCCGCAATTCGTTGTT 15051
| 3181 GAGTTTGGTTCGATTCCTGAGCTGCTTGAAGACGCTGGAACCCGACCGGCGGACCTTCCC 3240
| 15052 GAGTTTGGTTCGATTCCTGAGCTGCTTGAAGACGCTGGAACCCGACCGGCGGACCTTCCC 15111
| 3241 AATATCCGTTATCAATCGGCGGACAGCTACCCGGTCAAACTCAGCGATATGCAAA 3300
| 15112 AATATCCGTTATCAATCGGCGGACAGCTACCCGGTCAAACTCAGCGATATGCAAA 15171
| 3301 TCCGCGCGCATGAGGCGCTGAGCGAGGCTTCTATTATTAACACCCGTTCCAGACCGT 3360
| 15172 TCCGCGCGCATGAGGCGCTGAGCGAGGCTTCTATTATTAACACCCGTTCCAGACCGT 15231
| 3361 TTGCTGAGACTCTCCGATACGCGGATCGAGCGCGAGCGGACGTTGGCAAACTTCCAGC 3420
| 15232 TTGCTGAGACTCTCCGATACGCGGATCGAGCGCGAGCGGACGTTGGCAAACTTCCAGC 15291
| 3421 GGTTCGATGAGAGCGGCGCTTCAACCTCGTTCGTGGGAGGATGAGCGCATCGAGTGC 3480
| 15292 GGTTCGATGAGAGCGGCGCTTCAACCTCGTTCGTGGGAGGATGAGCGCATCGAGTGC 15351
| 3481 CTGTATGATCGTCTGTCGAGAAATTTTCTGCTGAGAGCGCGGATATATGCGGACGTC 3540
| 15352 CTGTATGATCGTCTGTCGAGAAATTTTCTGCTGAGAGCGCGGATATATGCGGACGTC 15411
| 3541 CTGATGAGAGAGCGCGCTCTCTGCAACATCGCGCTGTCGCTGGGCGCAATTCATTTT 3600
| 15412 CTGATGAGAGAGCGCGCTCTCTGCAACATCGCGCTGTCGCTGGGCGCAATTCATTTT 15471
| 3601 GAACAGGTTGCGGCGGCTTCTGCAACCTCGGACGATTCGGAAGTTTACGTCACGCGCATGCTG 3660
| 15472 GAACAGGTTGCGGCGGCTTCTGCAACCTCGGACGATTCGGAAGTTTACGTCACGCGCATGCTG 15531
| 3661 GCGCGGCGGCTTGAACCCGCGGAGTTTCAAGTCTGTACGCTCGTCAAGATTCCTTACCG 3720
| 15532 GCGCGGCGGCTTGAACCCGCGGAGTTTCAAGTCTGTACGCTCGTCAAGATTCCTTACCG 15591
| 3721 AGCGCGCGCAACGAGCGCGGCGCGCTTCCCGCGCGGAGACGATTCGCCGATATGCTT 3780
| 15592 AGCGCGCGCAACGAGCGCGGCGCGCTTCCCGCGCGGAGACGATTCGCCGATATGCTT 15651
| 3781 CGCGACTTCTTGAAGCAAACTACCGGATGATAGTGTGCTTCACTTCTGAGAGCTC 3840
| 15652 CGCGACTTCTTGAAGCAAACTACCGGATGATAGTGTGCTTCACTTCTGAGAGCTC 15711
| 3841 GATGCGTTGCGCTGATCGTCAACGCGCAAGTGTGATGAGGCGCTGCGCGAGCGGAG 3900
| 15712 GATGCGTTGCGCTGATCGTCAACGCGCAAGTGTGATGAGGCGCTGCGCGAGCGGAG 15771
| 3901 GATGCTCTGCTCGCGCGGAGATTCGCGGCGCAACGCGGACCGCTTGAAGAGATC 3960
| 15772 GATGCTCTGCTCGCGCGGAGATTCGCGGCGCAACGCGGACCGCTTGAAGAGATC 15831
| 3961 CTGCTGCGGCTGATGAGGAGGCTCGGCGCTGAGAGGATGATGAGGCTTCCAGAGAGCTTC 4020
| 15832 CTGCTGCGGCTGATGAGGAGGCTCGGCGCTGAGAGGATGATGAGGCTTCCAGAGAGCTTC 15891
| 4021 GTTCGATCTTGTGTCGATCGATTCATCGTTCCGATGAGAGCGCTGTTGCGAAGAG 4080
| 15892 GTTCGATCTTGTGTCGATCGATTCATCGTTCCGATGAGAGCGCTGTTGCGAAGAG 15951
| 4081 CTGATGAGGAGATTCGCGCATACCGAGTTGTTCAATACCGAAGCTCGGCTGCTGGG 4140
| 15952 CTGATGAGGAGATTCGCGCATACCGAGTTGTTCAATACCGAAGCTCGGCTGCTGGG 16011
| 4141 TCCGCTTTCGCGGAGACTCGAGAGATTCAGATCAGCGCGGCAATGACGAGGACGAGTGC 4200
|
```

```
DB 16012 TCCGCTTTCGCGCGGAGCTCGAGAGATCTAGATACGCGCGGAAATGACGAGCGAGTGC 16071
| 4201 GAGGTTCCGCGCGAGAGGCGAGAGACTGCTTAA 4233
| 16072 GAGGTTCCGCGCGAGAGGCGAGAGACTGCTTAA 16104
|
| RESULT 7
| US-09-567-899-1
| ; Sequence 1, Application US/09567899
| ; Patent No. 6383787
| ; GENERAL INFORMATION:
| ; APPLICANT: Schupp, Thomas
| ; APPLICANT: Ligon, James
| ; APPLICANT: Molnar, Ietvan
| ; APPLICANT: Zirkle, Ross
| ; APPLICANT: Cyr, Devon
| ; APPLICANT: Goerlach, Joern
| ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
| ; FILE REFERENCE: 4-30582A
| ; CURRENT APPLICATION NUMBER: US/09/567,899
| ; CURRENT FILING DATE: 2000-05-10
| ; PRIOR APPLICATION NUMBER: 09/335,409
| ; PRIOR FILING DATE: 1999-06-17
| ; NUMBER OF SEQ ID NOS: 30
| ; SOFTWARE: Patentin Ver. 2.0
| ; SEQ ID NO 1
| ; LENGTH: 68750
| ; TYPE: DNA
| ; ORGANISM: Sorangium cellulosum
| ; US-09-567-899-1
|
| Query Match 100.0%; Score 4233; DB 4; Length 68750;
| Best Local Similarity 100.0%; Pred. No. 0;
| Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
|
| 1 ATGAGATCAATCACTTCTGAAGAGCTCGAGACGAGGATATCAAGTGGCGGCGAT 60
| 11872 ATGAGATCAATCACTTCTGAAGAGCTCGAGACGAGGATATCAAGTGGCGGCGAT 11931
|
| 61 GGGAGCGCTTCAGATACAGGCGCCCAAGAACGCTGAACCCGAACTGCTGCTGA 120
| 11932 GGGAGCGCTTCAGATACAGGCGCCCAAGAACGCTGAACCCGAACTGCTGCTGA 11991
|
| 121 ATCTCGAGCAAAAGCAAGATCTGACGATGCTCGTCAAGACTCCCGGCAAAATCC 180
| 11992 ATCTCGAGCAAAAGCAAGATCTGACGATGCTCGTCAAGACTCCCGGCAAAATCC 12051
|
| 181 ATCGTCCGCGCGCGAGCGGAGCGCTCGTTCCTTCAAGACATCCAAATCC 240
| 12052 ATCGTCCGCGCGCGAGCGGAGCGCTCGTTCCTTCAAGACATCCAAATCC 12111
|
| 241 TACTGCTGGGCGGAGAGAGCTTTACGTTCCGTCACAGACATCCAAATCC 300
| 12112 TACTGCTGGGCGGAGAGAGCTTTACGTTCCGTCACAGACATCCAAATCC 12171
|
| 301 TAGCATGTGATCTGAGCTGCGGAGGCTGAGACCGCGCTTTCGAAAGTCTGCGG 360
| 12172 TAGCATGTGATCTGAGCTGCGGAGGCTGAGACCGCGCTTTCGAAAGTCTGCGG 12231
|
| 361 CGGACGACATGCTTGGGCGGCAACAGCGTCCGCGCATGATGAGGATGAGGCTTAA 420
| 12232 CGGACGACATGCTTGGGCGGCAACAGCGTCCGCGCATGATGAGGATGAGGCTTAA 12291
|
| 421 GTTCGAGCGCGACATGAGATCATGATCTGCGGCGCTGACCGGAGCACGCGAAGCG 480
| 12292 GTTCGAGCGCGACATGAGATCATGATCTGCGGCGCTGACCGGAGCACGCGAAGCG 12351
|
| 481 AGGCTGCTGCTGTTGCGAGATGCGATGTCGACCGCGCATCTATGACACGAGGCGCTCCG 540
| 12352 AGGCTGCTGCTGTTGCGAGATGCGATGTCGACCGCGCATCTATGACACGAGGCGCTCCG 12411
|
| 541 CTCTATCAAGTGTGCGCGCTTGGCTGAGACGAGCGGCAAAACCGTCTGCTCAGATTC 600
|
```

Db 12412 CTCTATCAGTCGTGCGCGTTGCGGCTGAGACGAGCGGCAAAACCGTCTGTGCTAGATATC 12471  
Qy 601 GATCTCATTAACGTTGACCTTAGGAGCCTGTCCATCATCTTCAAGAGTCTGCTGAGCTTC 660  
Db 12472 GATCTCATTAACGTTGACCTTAGGAGCCTGTCCATCATCTTCAAGAGTCTGCTGAGCTTC 12531  
Qy 661 TACGAGATCCCGAGACCTCTCTCCCTGTCTGAGAGCTCTGATACCGGATTAATGATCTC 720  
Db 12532 TACGAGATCCCGAGACCTCTCTCCCTGTCTGAGAGCTCTGATACCGGATTAATGATCTC 12591  
Qy 721 GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATCAACGATCGATGATTACTGAAAGCGG 780  
Db 12592 GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATCAACGATCGATGATTACTGAAAGCGG 12651  
Qy 781 GCGATCCCGAGACCTCCACCTCCCGGAGAGCTTCCGATGAAAGCCGATCTTACCTTC 840  
Db 12652 GCGATCCCGAGACCTCCACCTCCCGGAGAGCTTCCGATGAAAGCCGATCTTACCTTC 12711  
Qy 841 AAGGATCCGCTTCCGACACAGAGGAAATGCTGCGCTCGGACTCTCTGGGCTGATTC 900  
Db 12712 AAGGATCCGCTTCCGACACAGAGGAAATGCTGCGCTCGGACTCTCTGGGCTGATTC 12771  
Qy 901 AAGCGGCTGTGCGGAGAGCGGCGCTGAACCCGACGCGGCTCATCTGCGATTTTCC 960  
Db 12772 AAGCGGCTGTGCGGAGAGCGGCGCTGAACCCGACGCGGCTCATCTGCGATTTTCC 12831  
Qy 961 GAGGATCCGCGCGCTGAGAGCGGAGCCCGGTTTACGCTCAATACATACGCTTCAAC 1020  
Db 12832 GAGGATCCGCGCGCTGAGAGCGGAGCCCGGTTTACGCTCAATACATACGCTTCAAC 12891  
Qy 1021 CGGCTCCCGCTCATCCGCGGCTGAACGATATCACCGGAGACTTCAAGCTCATGCTTC 1080  
Db 12892 CGGCTCCCGCTCATCCGCGGCTGAACGATATCACCGGAGACTTCAAGCTCATGCTTC 12951  
Qy 1081 CTGACATCGACACCACTCGCGACAAAGACTTGAACAGCGGCTTAAGCTTAACAG 1140  
Db 12952 CTGACATCGACACCACTCGCGACAAAGACTTGAACAGCGGCTTAAGCTTAACAG 13011  
Qy 1141 CAGCTGTGGGAAGCATGATGATCATCTGCGAGCTTAAGCGGTTATGAGTCTCAGCGAGAGCC 1200  
Db 13012 CAGCTGTGGGAAGCATGATGATCATCTGCGAGCTTAAGCGGTTATGAGTCTCAGCGAGAGCC 13071  
Qy 1201 GCCCGGCTCTGGGAGTCAACAGAGGCGCATTTGTTCCCGTGTGCTCAGCGAGCGCTT 1260  
Db 13072 GCCCGGCTCTGGGAGTCAACAGAGGCGCATTTGTTCCCGTGTGCTCAGCGAGCGCTT 13131  
Qy 1261 AACGCAAGTCTGTGTGTCACTCTGTGCAAGAGCTCGGAATCCGCTGTACACGAC 1320  
Db 13132 AACGCAAGTCTGTGTGTCACTCTGTGCAAGAGCTCGGAATCCGCTGTGTACACGAC 13191  
Qy 1321 ACGGAGATCTCTCAGCTGCTGTGATCATCAGCTCTACAGACAGATGGGAGCTTCTCTC 1380  
Db 13192 ACGGAGATCTCTCAGCTGCTGTGATCATCAGCTCTACAGACAGATGGGAGCTTCTCTC 13251  
Qy 1381 CTCGCTGGGAGCATGTCGACGAGATGTTCCCGCCGACCTTCTGAGAGACATGCTCGAA 1440  
Db 13252 CTCGCTGGGAGCATGTCGACGAGATGTTCCCGCCGACCTTCTGAGAGACATGCTCGAA 13311  
Qy 1441 GCGTACGTCGTTTTCTCCGCGGCTCATCTGAGAACCATGGGAGTGAACAGGTGCGCTGT 1500  
Db 13312 GCGTACGTCGTTTTCTCCGCGGCTCATCTGAGAACCATGGGAGTGAACAGGTGCGCTGT 13371  
Qy 1501 TCGCTTCGCGCTGCCCACTAGAAAGCGGCGAGAGCCGAAACGACAGCGCTGCTG 1560  
Db 13372 TCGCTTCGCGCTGCCCACTAGAAAGCGGCGAGAGCCGAAACGACAGCGCTGCTG 13431  
Qy 1561 ACGGAGATTAAGCTCAAGGCTGTTTCGCGGCGGAGTCAAGCAGCTGCCATGACGCTC 1620  
Db 13432 ACGGAGATTAAGCTCAAGGCTGTTTCGCGGCGGAGTCAAGCAGCTGCCATGACGCTC 13491  
Qy 1621 GCGCTGTGTGCGCGCGCAAGACGCTCACTGACGAAGACTTTTCGCGCTTTCGCGCGCA 1680

Db 13492 GCGGTGTGTGCGCGCGCAAGACGCTCATCTGACGAAGAGCTTTCGCGCTTCGCGCGCA 13551  
Qy 1681 CTGGCGCGCGGCTGCCGAGACAGGGGCAACGCCGAACATTTGTGCGGCTGTGATG 1740  
Db 13552 CTGGCGCGCGGCTGCCGAGACAGGGGCAAGGCCGAACATTTGTGCGGCTGTGATG 13611  
Qy 1741 GAGAAAGCTGGGAGCAGAGTGTGCGGCTTCTCCGCTGTCTGAGTCAAGCGCGCTTAC 1800  
Db 13612 GAGAAAGCTGGGAGCAGAGTGTGCGGCTTCTCCGCTGTCTGAGTCAAGCGCGCTTAC 13671  
Qy 1801 GTGCCGATGATGCGGACCTTACCGGCGAGAGTATCCATCTCTCTGATCATGTGAG 1860  
Db 13672 GTGCCGATGATGCGGACCTTACCGGCGAGAGTATCCATCTCTCTGATCATGTGAG 13731  
Qy 1861 GTAAAGCTGTGAGAGCGGACCTGCTGATGAGCAAGCTGTATGCGCGCGCGGATC 1920  
Db 13732 GTAAAGCTGTGAGAGCGGACCTGCTGATGAGCAAGCTGTATGCGCGCGCGGATC 13791  
Qy 1921 CAGCGGCTGCTCTGAGAGCGGAGCGGCTCGAAAGCGACAGCGACCAAGCTTCGATGATG 1980  
Db 13792 CAGCGGCTGCTCTGAGAGCGGAGCGGCTCGAAAGCGACAGCGACCAAGCTTCGATGATG 13851  
Qy 1981 CCCATTGAGACCTTGGATCTTGGCTATGTCTATCTACCTGCGGATCCACAGGGTTG 2040  
Db 13852 CCCATTGAGACCTTGGATCTTGGCTATGTCTATCTACCTGCGGATCCACAGGGTTG 13911  
Qy 2041 CCCAAGGGGAGTATGATCATCGGGGTCGCTCAACACATCTGGAATCAACAGAG 2100  
Db 13912 CCCAAGGGGAGTATGATCATCGGGGTCGCTCAACACATCTGGAATCAACAGAG 13971  
Qy 2101 CGCTTCAAAATAGGAGCCCGAGACAGGGTGTGCGCTCTCTCTGCTGAGCTTCAATCTC 2160  
Db 13972 CGCTTCAAAATAGGAGCCCGAGACAGGGTGTGCGCTCTCTCTGCTGAGCTTCAATCTC 14031  
Qy 2161 TCGGCTATGATGTGTGCGGATCTTGGCGGCGGAGTCAATGCTGCTGCGAGCGG 2220  
Db 14032 TCGGCTATGATGTGTGCGGATCTTGGCGGCGGAGTCAATGCTGCTGCGAGCGG 14091  
Qy 2221 TCCAAGGCGCGATCCGCGGCAATGGGCAAGTGTGACAGAGAGAGTGAACGCTG 2280  
Db 14092 TCCAAGGCGCGATCCGCGGCAATGGGCAAGTGTGACAGAGAGAGTGAACGCTG 14151  
Qy 2281 TGAACCTCGGTGCGGCGCTGATCGGATGCTCTGACGATTTTGAAGGCTCGGCCGAT 2340  
Db 14152 TGAACCTCGGTGCGGCGCTGATCGGATGCTCTGACGATTTTGAAGGCTCGGCCGAT 14211  
Qy 2341 TCGCTCGTAACTCTTGGCGCTTTCCTGCTGAGCGGCGCATGATCCGCTGAGGCTG 2400  
Db 14212 TCGCTCGTAACTCTTGGCGCTTTCCTGCTGAGCGGCGCATGATCCGCTGAGGCTG 14271  
Qy 2401 CTGGCGAGCTCCAGGCAATCAGGCGCGGCTGTCCGTGATCAAGCTGGGCGGCGCAC 2460  
Db 14272 CTGGCGAGCTCCAGGCAATCAGGCGCGGCTGTCCGTGATCAAGCTGGGCGGCGCAC 14331  
Qy 2461 GAAAGCTGATCTGTGCTCATCGGAGTACCCCGTGAAGAAAGTCACTATCTGCGGAGC 2520  
Db 14332 GAAAGCTGATCTGTGCTCATCGGAGTACCCCGTGAAGAAAGTCACTATCTGCGGAGC 14391  
Qy 2521 ATCCCTTAAGCGGCTGCGGCTGCGCAACGAGCTTTCAGCTGTGATGAGCGCTCGAA 2580  
Db 14392 ATCCCTTAAGCGGCTGCGGCTGCGCAACGAGCTTTCAGCTGTGATGAGCGCTCGAA 14451  
Qy 2581 CCGGCGCGGCTGTGGGTTCCGGGCACTCTACATTTGGCGGCGGCTGGCACTGGGC 2640  
Db 14452 CCGGCGCGGCTGTGGGTTCCGGGCACTCTACATTTGGCGGCGGCTGGCACTGGGC 14511  
Qy 2641 TACTGCGGATGAAGAAAGACGCGCAAGGCTTCTCTGTGACCCCGAGACCGGGAG 2700  
Db 14512 TACTGCGGATGAAGAAAGACGCGCAAGGCTTCTCTGTGACCCCGAGACCGGGAG 14571  
Qy 2701 GCGCTTCAAAAGCGGCGATCTGGGCGGCTACCTGCGCGATGGAACATGAGTTATG 2760  
Db 14572 GCGCTTCAAAAGCGGCGATCTGGGCGGCTACCTGCGCGATGGAACATGAGTTATG 14631

QY 2761 GGGCGTGAAGACCAACCAATCAAGCTTCGCGATACCGGCTTGAGCTCGGGGAAATCGAG 2820  
DB 14632 GGGCGTGAAGACCAACCAATCAAGCTTCGCGATACCGGCTTGAGCTCGGGGAAATCGAG 14691  
QY 2821 GAAACCTCAAGTCGATCGAAGCTACGCGAGCGCGGTGATTTGCGCCGCGGAAACGAG 2880  
DB 14692 GAAACCTCAAGTCGATCGAAGCTACGCGAGCGCGGTGATTTGCGCCGCGGAAACGAG 14751  
QY 2881 GCGGCAACAAAGCTCTTCTAGCTATGTGTCCCGAGGGGACACGAGAGACGCGCTGCC 2940  
DB 14752 GCGGCAACAAAGCTCTTCTAGCTATGTGTCCCGAGGGGACACGAGAGACGCGCTGCC 14811  
QY 2941 GAGCAGAGACCGGAGCTTCAAGACCGGAGCGGATGAGAGCGGAGACCAACGCGCGGAAAGG 3000  
DB 14812 GAGCAGAGACCGGAGCTTCAAGACCGGAGCGGATGAGAGCGGAGACCAACGCGCGGAAAGG 14871  
QY 3001 GACGCGTTGAGCAGACGGGAGAGAGGTGAGTTCAAGCTCGCTGACACGGAATCCGGAAG 3060  
DB 14872 GACGCGTTGAGCAGACGGGAGAGAGGTGAGTTCAAGCTCGCTGACACGGAATCCGGAAG 14931  
QY 3061 GACCTGAGACGAAAGCCCGTCTGATCTGACCCGCGAGAGATCCGCGGAGCGGGGCTTG 3120  
DB 14932 GACCTGAGACGAAAGCCCGTCTGATCTGACCCGCGAGAGATCCGCGGAGCGGGGCTTG 14991  
QY 3121 GACGTCTAGCGCGGTGCGCGTACGCTCCGAAAGCTTCCTTGAGGCGCCGATCCGTTTGT 3180  
DB 14992 GACGTCTAGCGCGGTGCGCGTACGCTCCGAAAGCTTCCTTGAGGCGCCGATCCGTTTGT 15051  
QY 3181 GAGTTTGTGATTCCTGAGCTCTGAGAGCGGTGAGAGCGCGGAGCGGACCTTCC 3240  
DB 15052 GAGTTTGTGATTCCTGAGCTCTGAGAGCGGTGAGAGCGCGGAGCGGACCTTCC 15111  
QY 3241 AAATTCGTTATCATCGCGGAGACGATCCGCGTCAAACTACCGGATGTCAA 3300  
DB 15112 AAATTCGTTATCATCGCGGAGACGATCCGCGTCAAACTACCGGATGTCAA 15171  
QY 3301 TCCGCGGATCGAGGCGGTGAGCGAGCGGCTTATATTAACACCGCTTCGAGACCGT 3360  
DB 15172 TCCGCGGATCGAGGCGGTGAGCGAGCGGCTTATATTAACACCGCTTCGAGACCGT 15231  
QY 3361 TTGCTGAAGCTCTCCGATCAGGAGATCGAGCGGAGCGGACCTTCGCGAAACTTCGAC 3420  
DB 15232 TTGCTGAAGCTCTCCGATCAGGAGATCGAGCGGAGCGGACCTTCGCGAAACTTCGAC 15291  
QY 3421 TTGCTGAAGCTCTCCGATCAGGAGATCGAGCGGAGCGGACCTTCGCGAAACTTCGAC 3480  
DB 15292 TTGCTGAAGCTCTCCGATCAGGAGATCGAGCGGAGCGGACCTTCGCGAAACTTCGAC 15351  
QY 3481 CTGTATGATCGATCGGAGAAATTTGCTGCTGAGGCGGAGATATATGCGGAGCTC 3540  
DB 15352 CTGTATGATCGATCGGAGAAATTTGCTGCTGAGGCGGAGATATATGCGGAGCTC 15411  
QY 3541 CTGTATGATCGATCGGAGAAATTTGCTGCTGAGGCGGAGATATATGCGGAGCTC 3600  
DB 15412 CTGTATGATCGATCGGAGAAATTTGCTGCTGAGGCGGAGATATATGCGGAGCTC 15471  
QY 3601 GAAACAGTTTCGCGCGTTCGACCTGCGACATTCGAGCTTTAAGTCAAGGATGCTG 3660  
DB 15472 GAAACAGTTTCGCGCGTTCGACCTGCGACATTCGAGCTTTAAGTCAAGGATGCTG 15531  
QY 3661 GCGGCGGAGTACCGCGGAGTTCGAGGCTGATACGCTGAGGATTCCTACCG 3720  
DB 15532 GCGGCGGAGTACCGCGGAGTTCGAGGCTGATACGCTGAGGATTCCTACCG 15591  
QY 3721 AGCGCGGACGACGCGCGCGCTCCCGCGCGGAGAGACTTCGCGGATGCTT 3780  
DB 15592 AGCGCGGACGACGCGCGCGCTCCCGCGCGGAGAGACTTCGCGGATGCTT 15651  
QY 3781 CGGAGCTTTGAGACCAAACTACCGGATGATGATGCTTCAAGTCTTGTGAGCTC 3840  
DB 15652 CGGAGCTTTGAGACCAAACTACCGGATGATGATGCTTCAAGTCTTGTGAGCTC 15711

QY 3841 GATGGTTGCGGCTGACGCTCAAGCGCAAGGTCGATCGTAAGGCGCTGCGGAGCGAAG 3900  
DB 15712 GATGGTTGCGGCTGACGCTCAAGCGCAAGGTCGATCGTAAGGCGCTGCGGAGCGAAG 15771  
QY 3901 GATACCTGCTCGCGCGGATTCGCGGAGCAACGCGGACCGGAGCGCTTGAGAGATC 3960  
DB 15772 GATACCTGCTCGCGCGGATTCGCGGAGCAACGCGGACCGGAGCGCTTGAGAGATC 15831  
QY 3961 CTGCTGCGGCTGATACGAGAGTCTCGGCGCTGAGAGTGTGCGGCTCAAGAGCTTC 4020  
DB 15832 CTGCTGCGGCTGATACGAGAGTCTCGGCGCTGAGAGTGTGCGGCTCAAGAGCTTC 15891  
QY 4021 GTGATCTTGTGAGCAATCGATTCATCGTTCGATGAGAGGCTGTCAGAGAGG 4080  
DB 15892 GTGATCTTGTGAGCAATCGATTCATCGTTCGATGAGAGGCTGTCAGAGAGG 15951  
QY 4081 CTGATAGGAGATCGCATCAACCGAGTTGTCAGTACCGAACCTCGGCTCGTGCGG 4140  
DB 15952 CTGATAGGAGATCGCATCAACCGAGTTGTCAGTACCGAACCTCGGCTCGTGCGG 16011  
QY 4141 TCCGTTTCCGCGGAGATTCGAGAGATCTAGATCAGCGCGCGGACATGACGAGCTG 4200  
DB 16012 TCCGTTTCCGCGGAGATTCGAGAGATCTAGATCAGCGCGCGGACATGACGAGCTG 16071  
QY 4201 GAGTTCCGCGGAGAGGCGAGAGCTGACTAA 4233  
DB 16072 GAGTTCCGCGGAGAGGCGAGAGCTGACTAA 16104

RESULT 8  
US-09-443-501A-2  
Sequence 2, Application US/09443501A  
Patent No. 630342  
GENERAL INFORMATION:  
APPLICANT: Kosan Biosciences, Inc.  
APPLICANT: Julien, Bryan  
APPLICANT: Katz, Leonard  
APPLICANT: Khosla, Chaitan  
APPLICANT: Tang, Li  
APPLICANT: Ziermann, Rainer  
TITLE OF INVENTION: Recombinant Methods and Materials for Producing  
TITLE OF INVENTION: Epochlone and Epochlone Derivatives  
FILE REFERENCE: 30062-20031.00  
CURRENT APPLICATION NUMBER: US/09/443,501A  
CURRENT FILING DATE: 1999-11-19  
PRIOR APPLICATION NUMBER: US 60/130,560  
PRIOR FILING DATE: 1999-04-22  
PRIOR APPLICATION NUMBER: US 60/122,620  
PRIOR FILING DATE: 1999-03-03  
PRIOR APPLICATION NUMBER: US 60/119,386  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: US 60/109,401  
PRIOR FILING DATE: 1998-11-20  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 71989  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic construct  
US-09-443-501A-2

Query Match 97.2%; Score 4114.6; DB 4; Length 71989;  
Best Local Similarity 98.3%; Pred. No. 0;  
Matches 4159; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGAGCATTAATCAGTTCTGAAAGAGCTCGAGCAACGAGGATCAAGCTGCGGCGCAT 60  
DB 6260 ATGAGCATTAATCAGTTCTGAAAGAGCTCGAGCAACGAGGATCAAGCTGCGGCGCAT 6319  
QY 61 GGGAGCGGCTCAATACAGGCGCCCAAGAGCGGCTGAAACCGAAGCTGCTGCTCGA 120

Dh 6320 GGGGAGCCCTCCAGATACAGGCCCCCAAGAACGCTTGAAACCCGAACTGCTCGCTGA 6379  
Qy 121 ATCTCCGAGCAACAAAGACGATCTCTGACGATGCTCCGTCAGAGACTCCCGGACAGATCC 180  
Db 6380 ATCTCCGAGCAACAAAGACGATCTCTGACGATGCTCCGTCAGAGACTCCCGGACAGATCC 6439  
Qy 181 ATCTGTCGCGCCCGGAGCGGAGCGCTCCGTTCTCTTCAAGACATCCAGAAATCC 240  
Db 6440 ATCTGTCGCGCCCGGAGCGGAGCGCTCCGTTCTCTTCAAGACATCCAGAAATCC 6499  
Qy 241 TACTGTCGCGCGGAGCGGAGCGCTTAAAGTCCCGAGCGGATCCAGCGCTATGCGGA 300  
Db 6500 TACTGTCGCGCGGAGCGGAGCGCTTAAAGTCCCGAGCGGATCCAGCGCTATGCGGA 6559  
Qy 301 TACGACTGTACGAGATCTGACGCTGCGAGGCTGAGCCGCGCTTTCGGAAGTCTGCGG 360  
Db 6560 TACGACTGTACGAGATCTGACGCTGCGAGGCTGAGCCGCGCTTTCGGAAGTCTGCGG 6619  
Qy 361 CGGACGACATGCTTCCGCGCCCAACGCTGCGCCGACATGATGCAAGTATGAGCCTTAA 420  
Db 6620 CGGACGACATGCTTCCGCGCCCAACGCTGCGCCGACATGATGCAAGTATGAGCCTTAA 6679  
Qy 421 GTCCAGCGCGGACATGATGATCATCTGCGCGCGGCTGACCGGAGCAACCGGAGCG 480  
Db 6680 GTCCAGCGCGGACATGATGATCATCTGCGCGCGGCTGACCGGAGCAACCGGAGCG 6739  
Qy 481 AGGCTGATGCTGTCGAGATGCGATGTCGACCGGATCTTATGACCGGAGCGGCTTCG 540  
Db 6740 AGGCTGATGCTGTCGAGATGCGATGTCGACCGGATCTTATGACCGGAGCGGCTTCG 6799  
Qy 541 CTCTATCACTGTCGCGCTTCCGCTGAGACGAGCGGAAACCGCTCTGCTGCTCAATTC 600  
Db 6800 CTCTATCACTGTCGCGCTTCCGCTGAGACGAGCGGAAACCGCTCTGCTGCTCAATTC 6859  
Qy 601 GATCTCATTTAAGTGTGACCTTGAAGCGCTGTCATCATCTTGAAGACTGCTGCTTC 660  
Db 6860 GATCTCATTTAAGTGTGACCTTGAAGCGCTGTCATCATCTTGAAGACTGCTGCTTC 6919  
Qy 661 TACGAGATCCCGAGACCTCTCCCTGCTGAGGCTCTGCTGAGGATGATGATGATGATC 720  
Db 6920 TACGAGATCCCGAGACCTCTCCCTGCTGAGGCTCTGCTGAGGATGATGATGATGATC 6979  
Qy 721 GCGCTGAGTCTCGCAAGAGTCTGAGGCGGATCAACGATCGATGATGATGATGATGATC 780  
Db 6980 GCGCTGAGTCTCGCAAGAGTCTGAGGCGGATCAACGATCGATGATGATGATGATGATC 7039  
Qy 781 CGCATCCCGAGCTCCACTCCCGGACGCTTCCGATGAAAGCCGATCATCACTTCG 840  
Db 7040 CGCATCCCGAGCTCCACTCCCGGACGCTTCCGATGAAAGCCGATCATCACTTCG 7099  
Qy 841 AAGGAGATCCCGCTCCGAGCAACGAGGATGCTGCGGATCCCTGGGCTGATGATG 900  
Db 7100 AAGGAGATCCCGCTCCGAGCAACGAGGATGCTGCGGATCCCTGGGCTGATGATGATG 7159  
Qy 901 AAGCGGCTGTCGAGGAGCGGCGGCTGACCGCGAGCGGATCTGCTGCTGATTTTC 960  
Db 7160 AAGCGGCTGTCGAGGAGCGGCGGCTGACCGCGAGCGGATCTGCTGCTGATTTTC 7219  
Qy 961 GAGGTGATCGGCGCTGAGGCGGAGCGGCTTTCGCTCAATCAATCAATCAATCAATCA 1020  
Db 7220 GAGGTGATCGGCGCTGAGGCGGAGCGGCTTTCGCTCAATCAATCAATCAATCAATCA 7279  
Qy 1021 CGGCTCCCGCTCATCCGCGGCTGAGGATGATCAACGAGGATCTTCACTGATGATGATC 1080  
Db 7280 CGGCTCCCGCTCATCCGCGGCTGAGGATGATCAACGAGGATCTTCACTGATGATGATC 7339  
Qy 1081 CTGACATCGACACCACTTCGAGCAAGAGCTTTCGAACAGCGGCTTTCGAAGAG 1140  
Db 7340 CTGACATCGACACCACTTCGAGCAAGAGCTTTCGAACAGCGGCTTTCGAAGAG 7399  
Qy 1141 CAGCTGTGGAGAGGATGATCATCTGAGCAAGGATGATGATGATGATGATGATGATGATG 1200  
Db 7400 CAGCTGTGGAGAGGATGATCATCTGAGCAAGGATGATGATGATGATGATGATGATGATG 7459

Qy 1201 GCCCGGCTCTGGGATCCAAAGAGCGCATTTGTTCCCGGTGCTCAAGAGCGGCTT 1260  
Db 7460 GCCCGGCTCTGGGATCCAAAGAGCGCATTTGTTCCCGGTGCTCAAGAGCGGCTT 7519  
Qy 1261 AACGAGAGTCTGTTGCTGCTCTGTCAGAGGCTCGGAACCTCGGTATCAACAGC 1320  
Db 7520 AACGAGAGTCTGTTGCTGCTCTGTCAGAGGCTCGGAACCTCGGTATCAACAGC 7579  
Qy 1321 ACGGAGACTCTGAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
Db 7580 ACGGAGACTCTGAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7639  
Qy 1381 CTCGCTGGGACATCTGTCAGAGAGTGTTCGCGCCGACCTTCTGAGACGATGCTGAA 1440  
Db 7640 CTCGCTGGGACATCTGTCAGAGAGTGTTCGCGCCGACCTTCTGAGACGATGCTGAA 7699  
Qy 1441 GCGTACGCTGTTTTTCTCCGCGGCTCACTGAGAAACATGAGGATGAAAGAGTGCCTGT 1500  
Db 7700 GCGTACGCTGTTTTTCTCCGCGGCTCACTGAGAAACATGAGGATGAAAGAGTGCCTGT 7759  
Qy 1501 TCGCTTCGCTGCTGCGCTGAGAGCGCGGCGAGCGGAAACGAGCAACGAGCGCTGCTG 1560  
Db 7760 TCGCTTCGCTGCTGCGCTGAGAGCGCGGCGAGCGGAAACGAGCAACGAGCGCTGCTG 7819  
Qy 1561 AGCGAGCATACGCTGCAAGGCTGTTCCGCGGCGGCGGCTGAGAGCTGCCATGAGCTTC 1620  
Db 7820 AGCGAGCATACGCTGCAAGGCTGTTCCGCGGCGGCGGCTGAGAGCTGCCATGAGCTTC 7879  
Qy 1621 GCGGTGCTGTCGCGCGGAGAGCGCTCACTGACAGAGCTTTTCGCGCTTCGCGGCA 1680  
Db 7880 GCGGTGCTGTCGCGCGGAGAGCGCTCACTGACAGAGCTTTTCGCGCTTCGCGGCA 7939  
Qy 1681 CTTGGCGCGCGGCTGCGGAGAGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
Db 7940 CTTGGCGCGCGGCTGCGGAGAGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 7999  
Qy 1741 GAGAAAGCTGGAG 1800  
Db 8000 GAGAAAGCTGGAG 8059  
Qy 1801 GTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
Db 8060 GTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8119  
Qy 1861 GTAAAGCTCTGCTGAG 1920  
Db 8120 GTAAAGCTCTGCTGAG 8179  
Qy 1921 CAGGCGCTGCTGAG 1980  
Db 8180 CAGGCGCTGCTGAG 8239  
Qy 1981 CCAATTCAGACACCTTCGATCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2040  
Db 8240 CCAATTCAGACACCTTCGATCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 8299  
Qy 2041 CCAAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100  
Db 8300 CCAAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8359  
Qy 2101 CGCTTCGAATGAGGAGCGGAG 2160  
Db 8360 CGCTTCGAATGAGGAGCGGAG 8419  
Qy 2161 TCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220  
Db 8420 TCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8479  
Qy 2221 TCAAGCTGCGGATTCGCGGAG 2280  
Db 8480 TCAAGCTGCGGATTCGCGGAG 8539



2281 TGGAACTGCTGCGCGCTGATGCGGATGCTGTGAGACATTTTGGAGGTCGCCCGCAT 2340  
DB |||||  
8540 TGAATCTGCTGCGCGCTGATGCGGATGCTGTGAGACATTTTGGAGGTCGCCCGCAT 8599  
QY 2341 TGGCTGCTGAGGCTCTTGGCGCTTTGCTGCTGAGCGCGACTGATCCCGGTGGGCTG 2400  
DB TGGCTGCTGAGGCTCTTGGCGCTTTGCTGCTGAGCGCGACTGATCCCGGTGGGCTG 8659  
QY 2401 CCTGCGAGCTCCAGGCTCATGAGCGCGCGGTGTGCTGATGATGCTGGGCGGCTCAC 2460  
DB CCTGCGAGCTCCAGGCTCATGAGCGCGCGGTGTGCTGATGATGCTGGGCGGCTCAC 8719  
QY 2461 GAAAGCTGATCTGCTGCTGAGGATGCGCGGATGCTGAGGATGCTGAGGATGCTGAG 2520  
DB GAAAGCTGATCTGCTGCTGAGGATGCGCGGATGCTGAGGATGCTGAGGATGCTGAG 8779  
QY 2521 ATCCCTTACGCGCTGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAG 2580  
DB ATCCCTTACGCGCTGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAG 8839  
QY 2581 CCGCGCTGCTGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 2640  
DB CCGCGCTGCTGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 8899  
QY 2641 TACTGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 2700  
DB TACTGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 8959  
QY 2701 CCGCTTACGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 2760  
DB CCGCTTACGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 9019  
QY 2761 GGGCGTGAAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 2820  
DB GGGCGTGAAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 9079  
QY 2821 GAAAGCTGATGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 2880  
DB GAAAGCTGATGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 9139  
QY 2881 GCGCGTGAAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 2940  
DB GCGCGTGAAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 9199  
QY 2941 GAGCGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3000  
DB GAGCGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9259  
QY 3001 GAGCGTGAAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 3060  
DB GAGCGTGAAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 9319  
QY 3061 GAGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3120  
DB GAGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9379  
QY 3121 GAGCTTACGCGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 3180  
DB GAGCTTACGCGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 9439  
QY 3181 GAGTTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 3240  
DB GAGTTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 9499  
QY 3241 AAATTCCTTATCATCGGCGGAGACGATCCCGGTGAAACCTACGCGATGTCGAA 3300  
DB AAATTCCTTATCATCGGCGGAGACGATCCCGGTGAAACCTACGCGATGTCGAA 9559  
QY 3301 TCGCGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3360  
DB TCGCGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9619  
QY 3361 TTGCTGAGGATGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3420

DB 9620 TTGCTGAGGATGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9679  
QY 3421 GTTTCGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3480  
DB GTTTCGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9739  
QY 3481 CTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3540  
DB CTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9799  
QY 3541 CTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3600  
DB CTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9859  
QY 3601 GAAAGCTTGGCGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 3660  
DB GAAAGCTTGGCGCTGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGATGCTGAGGAT 9919  
QY 3661 GCGCGGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3720  
DB GCGCGGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 9979  
QY 3721 AGGCGGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3780  
DB AGGCGGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 10039  
QY 3781 CCGGCTTCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 3840  
DB CCGGCTTCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 10099  
QY 3841 GATGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 3900  
DB GATGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 10159  
QY 3901 GATGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 3960  
DB GATGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 10219  
QY 3961 CTGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 4020  
DB CTGCTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 10279  
QY 4021 GTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 4080  
DB GTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 10339  
QY 4081 CTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 4140  
DB CTGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 10399  
QY 4141 TCGGCTTGGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 4200  
DB TCGGCTTGGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 10459  
QY 4201 GAGGCTTGGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 4233  
DB GAGGCTTGGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 10492

RESULT 9  
US-09-252-991A-4765  
; Sequence 4765, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.116  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18





```

Db      2053  CGCCCGCGCGAGGTGGCGGCGAGCACTGCGCTGATGATCTACACCTTCGGCTCC 2112
Qy      2032  ACAGGGTTCCCAAGGGGGTGTATGATCATCGGGGTGCGGTCAACACATCTCGAAC 2091
Db      2113  ACCGGACGCCAAGAGGGCGGTGATGCTCAGGCATGCGGCGGTGAGCAACGCGTCTGAC 2172
Qy      2092  ATCAACGAGCGCTTGCATAATAGGGCCCGAGAGAGGGTGTGCGCTCTCTCTGCTGAC 2151
Db      2173  ATCAACACGCGCTACGCGCTGACGCGCAACGACCGGCTCTCTGCGCTCGAGCTGAC 2232
Qy      2152  TTGCATCTCTCGCTATGATGATGTTGCGGATCTCTGCGGCGGGCGGTACGATCGTG 2211
Db      2233  TTGCACCTCTCGGTCTACGATCTTCTGCGGCGCAACGCGGCGGGGCCACGATGCTCT 2292
Qy      2212  CCGGACGCGCTCAAGCTGCGCGATTCGCGCATTTGGCGAGATTGATCAACGAGAGAA 2271
Db      2293  CCGGACCGCGCGCGCGAGCGCATTCATCGCACTGGCGGAACTGCTGGAAACGCCAGCC 2352
Qy      2272  GTGACGGTGTGAACTCGGTGCGCGGCTGATGCGGATGCTCGTCAAGCATTTTGAAGGT 2331
Db      2353  ATCACCTGTGAACTCGGTGCGCGCGCAAGCGCATATCTCATATTAACCTGAGAAC 2412
Qy      2332  CGCGCGGATTCGCTGCTAGTGTCTCTGCGGCTTTCGCTGAGCGGCGCATGATCCCG 2391
Db      2413  GAGCGCGCAACGTACCTGCGCGGAGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2472
Qy      2392  GTGGGCGTCTGCGCGAGCTCCAGGCATCAAGCGCGCGGTGCTGCTGCTGCTGCTGCT 2451
Db      2473  GTGACGCTGCGCAACCGGATGATGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCT 2532
Qy      2452  GGGGCGCACCGAAGCGTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2511
Db      2533  GCGCGCACCGAAGCGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2592
Qy      2512  TGGGCGGAGTCCCTCAACGCGCGCTGCGCGCAACGCGCTGCTGCTGCTGCTGCTGCT 2571
Db      2593  CTGGCGCGGATCTCTTATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2652
Qy      2572  GCGCTCGAACCGCGCGGTCTGCGGTTCGCGGCGCACTTCACTTGGCGGCGGTGCGG 2631
Db      2653  CCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2712
Qy      2632  GCACTGGGTCTCTGCGCGGATGAAAGAGACCGCGGAGAGCTTCTCTGCTGACCGCG 2691
Db      2713  GCGCTCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2771
Qy      2692  ACCGCGGAGCGCTCTCAACAGACCGCGCGATCTGCGCGCTGCTGCTGCTGCTGCTG 2751
Db      2772  --TGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2829
Qy      2752  GAGTTCATGCGCGGTGAGAGCAACCAATCAAGCTTTCGCGGATACCGCGTTCGCGG 2811
Db      2830  GAGTTCCTGCGCGCGCGAGAGCAACCAAGTTCGCGGCGCGCGCGCGCGCGCGCGCG 2889
Qy      2812  GAAATGAGAGAAACGCTCAAGTTCGCGAGTTCGCGGAGCGCGGTGATTCGCGCGCT 2871
Db      2890  GAGCTGAGAGCGCGCGTGTGCGCTCAATTCGCGAGTTCGCGGCGCGCGCGCGCGCT 2949
Qy      2872  GGGAGCAGCGCGCGCGAACAAGCT 2894
Db      2950  GCGGAGACCCAGAGCGCGAGCT 2972

```

```

RESULT 10
US-09-252-991A-4675/c
; Sequence 4675, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

```

```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; . CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4675
; LENGTH: 2865
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4675

Query Match      11.6%; Score 491; DB 4; Length 2865;
Best Local Similarity 54.8%; Pred. No. 9,9e-99;
Matches 1121; Conservative 0; Mismatches 885; Indels 39; Gaps 6;

Qy      850  CGCTTCGCGCACACGAGCAATGCTGCGGTGAGACTCTGCGGCTGATGATGAGCGGCT 909
Db      2811  CGCTTCGCGCACTTCCACGCGCGCTTCAGACAGAGCGCTGCGAGCGCTGCGCGCGCG 2752
Qy      910  GTGGGAGAGCGGCGCTGACCCGACGCGGCTGATCTGCTGCAATTTTCGAGGTATC 969
Db      2751  GCGGCGAACAACGCGCTGAGCGCTGCGCGGCTGCTGCGCGCTTGGCGGCTTGGCGGACATC 2692
Qy      970  GGGCGCTGAGCGCGACGCGCGCGCTTACGCTCAACATTAAGCTCTTCAACCGGCTCCC 1029
Db      2691  GGTGCTGTGAGCAGGACCGCGCGCTTGTCTCAACCTGACGAGTACTCAACCGGCGCG 2632
Qy      1030  GTCCATTCGCGCGGTGAAGATATCAACCGGCGCTTCAAGTGTGCTCTCTGAGATC 1089
Db      2631  CTGCATTCGCGAGCTGCGCGAGTGTCTGCTGATCTTCAACGCGCTGCTGCGAGT 2572
Qy      1090  GACACCACTCGGCAAGAGCTTTCGAACAGCGCGCTTAAAGCTATTCAAGAGAGCTGTG 1149
Db      2571  GACACCGCGCACGCGGAGAGATTTCTGCGAGCGTGCACCGCATGCGGAGAGAGATGTTTC 2512
Qy      1150  GAAAGATGATCACTGCGAGCTTAAGCGGTATTCGAGGTTCAGCGGAGAGCGCGCGGCTC 1209
Db      2511  GACGACCTGACACCGACCGACTTCAGCGCGCGCTGACCTGCTGCGGAACTGCGCGG---C 2455
Qy      1210  CTGGGAGATCAAGAGCGCATTTGTTCCCGGTGTCTCACAGAGCGCGCTTAACAGCA 1269
Db      2454  CCGCGCGGTGCGGCGCGCGAGCTGATGCGCGGTGTTCACAGTGGATCGGAGCGG 2395
Qy      1270  GTGTTGTGTCACTCTGTCAGAGGCTCGAACTCCGCTGTATCAACAGCAGCACT 1329
Db      2394  CAGCGCTGCTGCGGATGAGCGAGCGCGCGCGCGCGCGCATATGATGACGAGACC 2335
Qy      1330  CCTGAGCTGCTGCGGATGATGAGCTTACAGAGCAGATGGGGAACCTGCTCGCGTGG 1389
Db      2334  CCGGAGTCTGAGTGAATGCTGAGGTACCGACGAGTTCGCGCGCTGAGATCGGCTGG 2275
Qy      1390  GACATGTCGACGAGAGTGTTCGCGCGCACTTCTGAGAGCAATCTCGAACCGTACGTC 1449
Db      2274  GACGTAAAGCTCTGAGTGTTCGCGAGGCGAGGAGGAGCATGTTGACGACTTCGTC 2215
Qy      1450  GTTTTCTTCGCGCGCTCACTGAGAACCATGAGGAGTGAACAGTGCCTGTTGCTTCG 1509
Db      2214  GGGCTGCTCGCGCGCTGCGCGAGCGCGCGCGCTGAGACCGAGCGGATGCCAGGA 2155
Qy      1510  CCGCGCGAGTAAAGCGCGCGCGAGCGGAACCGGACCAACGCGCTGCTGAGCAGAT 1569
Db      2154  CCGGTGAGGCGCGCGCGAGCGGT-----GCGCGTATGTCCTCCG 2113
Qy      1570  AGCTGCAAGCGCTTTCGCGCGCGGCTGAGAGCTGCGCATGCTGCGCTGCTGCTGCT 1629
Db      2112  AGCATGCGCGCGGTTTTCGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2053
Qy      1630  TCGCGCGCAGAGCGCTTACGTAAGAGCTTTTTCGCGCGCTTTCGCGCGAGCTTTCGCG 1689
Db      2052  GATGCGCGCGGAGTACGATACGATACCGGAGGTGCGCGCGAGCAGCAGCGCGCTGCGCG 1993

```

Qy	1690	CGCGTCGCGCAGACGAGGGGGGACACCGCCGAAACATATGGCTCGCGGTGTATGAGAAAGGC	1749
Db	1992	GTCTTGAAAGCGCACGAGCGGGGCGCTGTGGCGGGGTGGGGATCTCTGCGAAAGGC	
Qy	1750	TGGAGACAGATGTGTCTGCGCGTTCTCGCGGTGTCTGAGTAGAGCGCGGCTACGTGCGCATC	1809
Db	1992	GCGGCGGAATTGTGTGGGGTGTATGGGATCTCTCAAGGCGGGGCGCGCTTATGTGCGGTG	1873
Qy	1810	GATGCCGACCTTACCGGGCGGAGCGTATCCACTACTCTTCGATCATGTGAGGTAAAGCTC	1865
Db	1872	GACATCCCGCCAGCCCTCCGCTGGGGGCCAGGAGATCTTCGCGAGCGCCGA-----A	1822
Qy	1870	GTGCTGACGACGACCATTGGCTGGATATGGCAAACGTATATGGCCGCGGGGATCCAGCGCTG	1929
Db	1821	GTGGTGGCGCTGTGTCTCTGGAAGACA--TGTCGCGAGGTCTGGCTGCGCTGCGTGG	1764
Qy	1930	CTCGTGACGAGCGAGCGGCGGTGAGAGGCGCACGCGCTCCAGTATGATGCCCATTTGAG	1989
Db	1763	CCATCGACCGGCGTG----GCGCGGACAGCGCGTGTGGCGCGCACCGCGCGCGGAGGTG	1708
Qy	1990	ACACCTTCGATCTTCGCGTATGTATCTTACACTTCGCGGATTCACAGGGTTGCCAAGGGG	2049
Db	1707	GCGGCGGACGACCTCGACTTAAGTATCTTACACTTCGCGGCTCCACCGGACGCGAAAGGGC	1648
Qy	2050	GTGATGATTCATTCGCGGGGTGCGGTCAACACATCTTGAGCATATAAGAGCGTTTCGA	2109
Db	1647	GTGATGTCTGACCACTGCGGCGGTGAGCAACAGCGTCTGCAATCTACCAAGCGCTACCGC	1588
Qy	2110	ATTAGGGCCCGGAGACAGGGGTGCTGGCGCTCTCTCGTGAAGCTTGATCTCTCGGCTAT	2165
Db	1587	GTCCAGCGCAACGACCGGCTCTCTCGGCTCTCGGCTCGGCGAGCTGAGCTTTCGATCTTCTG	1528
Qy	2170	GATGTGTTTGGGATCTCTGCGGCGGGGCGGTACGATGTGTGTCCGGAACGCTTCMAAGCTG	2229
Db	1527	GACTTCTTCGCGCGCACCGCGGCGGGGGCCAGGGGTGCTCTCCGGAACCGGCGCGGCGC	1468
Qy	2230	CGCGATCCGGCGCATTTGGGCGAGATTTACAGAGAGAGAGATGACGCTGTGAACTCG	2288
Db	1467	AGCATTCATGTGCACTGGGCGGAACTGTGAAACCCACGCCATCACCTGTGAAACTCG	1408
Qy	2290	GTGCGGCGGCTGATGCGGATGCTGCTGAGCACTTTGAGGGGTGCGCCGATTCGCTGCT	2349
Db	1407	GTGCGGCGCCAAAGCGCAAGATGCTACGATTACTTGAGAGAGAGACCGCAACGTACCTG	1348
Qy	2350	AGGTCTCTGCGGCTTTTCGCTGTCGAGCGGCGACTGGAATCCGGTGGGCGCTCTGGCGAG	2409
Db	1347	CGGGGACCGGCGCTGTGTGCTCTGTGTCCGGTACTGGAATTCGGGTACGCTGCCAACCGCA	1288
Qy	2410	CTCCAGGCGCATACGCGCGCGGTGTGCGGTATACGCTGGGCGGGGCGCAACGAAGCTCG	2465
Db	1287	TGTGTGGCGGCGCTGTGCGGACGACGCGCTGTTCAGCGCTGGGCGGCGCGCACCGAGCGGCG	1228
Qy	2470	ATCTGATCATTCGGGTATCCCGGTAGAGAAAGTCCGACCTATGTGGGCGAGCATCCCTTAC	2529
Db	1227	ATCTGTGTGATTCGAGCAAGCCCATTCGCGCGGACGACACCGAGCTGGCGCAGATCTTAT	1168
Qy	2530	GCGCGTCCGCTGCGGACACGACGTTCCACGTCGCTGATGAGCGGCTCGAACCAGCGCGCG	2589
Db	1167	GCGCGTGGCTCTGCGGCGGACAGAGCTGGAATGCTCGATGTCCCGGGGCGGCGCTGCGCG	1108
Qy	2590	GTCGTGGTTTCGGGGCAACTCTACATTGGGCGGGGTGGGCTTGGCATGAGGCTTACGCGC	2649
Db	1107	CCGGGCGCGGCGGCGGAGATCCATTCATTCGGCGGGGTGGGCTTGGCGCTCGGCTACCGCGGC	1048
Qy	2650	GATGAAAGAAAGACCGCGAAGAGCTTCTCTGTGCAACCCCGAGACCGGGAGCGCTTAC	2709
Db	1047	GATCCGACGCGCACCGCGAAAGCTTGTGTCTGCTACCCCGA--TGCGCTGTGCTGTAT	991
Qy	2710	AAGACCGGCGATTTGGGCGCGCTACCTGCGCGCATGAGAAACATGAGATTATGAGGGCGTAG	2765
Db	990	CGACACGCGGACCTTCGCGCGCTATTTGGCGCAGCGGACAGATCGAATTCCTTCGCGCGGAG	931
Qy	2770	GACAAACAAATCAAGCTTCGCGGATACCGCGTTGAGCTTCGGGGAATTCAGAGAAACGCTC	2829

```

Db      930 GACACCAAGGTAAGATTCCGGCCACCCGATCGAATGGCCGACCTGACGCCGCGCTG 871
QY      2830 AAGTCGATCCGAAGTACGACGCGGTGATGTGCGCCGTGGGAAAGACGCGCGCAAC 2889
Db      870 TGGCATCATCCGACGATCAATCTGGCGGCCACCGTGTGCTCGGCGAGACCAACGACGCG 811
QY      2890 AAGCT 2894
Db      810 AGCCT 806

RESULT 11
US-09-252-991A-4815
; Sequence 4815, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4815
; LENGTH: 2922
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4815

Query Match      10.7%; Score 455; DB 4; Length 2922;
Best Local Similarity 50.1%; Pred. No. 7,9e-91;
Matches 1352; Conservative 0; Mismatches 1295; Indels 50; Gaps 7;

QY      156 CCGTCAGAGACTCCCGCAGAAATCCATCGTCCGCCGCCAGCCGAGCGGACGCTCCGT 215
Db      264 CCGGCTTCGCGCCCGGCGCAAGGTGCGCTGCCGAGCGGCGCAGAGATGGCATGACCGCTT 323
QY      216 TCCTCTCAGACATCCAGAAATCTTACTGCGTGGGCC---GGACAGGAGCGTTTACGT 272
Db      324 CGACCTCTCTCCGTCGACGAGGCTCTCTGCGTGGGACGTGGCGCCGCGCAGAGTGTGGG 383
QY      273 CCCGAGCGGATCCAGCGCTATCGCGAAATCGAATCGATCGATCTGACGTGCGAGGCT 332
Db      384 CAACTGACTGCGCCATCTCTCTGGAATTCGCGCGCGGAGATGCCACCGCAGCGCT 443
QY      333 GAGCGCGGCTTTGGAAATCGTCCGCGGCGACGACGATCTTCGGGCCACGCGTGC 392
Db      444 GCGCGCGGCGGCGAGTGGCTGCTCAACCGCACCGATGTTGGGGCGGGCTTCCGCA 503
QY      393 CGACATGATGAGGTGATCGAGCTTAAGTCGACGCCGACATGAGATCATCATCTGCG 452
Db      504 CGGTGCGCAGAGATCTTCCGACGCGCGCTGTCTCTGCTTCGACCTGACGAGACTGGCG 563
QY      453 CGGGCTGACCCGAGACACAGGAGAGGAGCTGTCTGTTCCGAGATCCGATGTGCA 512
Db      564 CACCTTCAGGTGAGCAGGCGCGGACGCGCATGCGCAGGCGCTGCGCAGCTGGCGCGCA 623
QY      513 CCGCATGTATGACACGAGCGGCGCTTCGCTCTATACGATGATGCGCGCTTGGCTGACGA 572
Db      624 TGAATGCTGCGGTGAGCGCGCGCGCAGGTGTTCTCTCTGGGCTGTGTCGATCCGCG 683
QY      573 GCGGCAAAACCGTCTCGTGTCAAGATCATATTAAAGTTGACCTAGCAGCGCTGTC 632
Db      684 CCGGAGGATTCGCTTGGCTGAGTCTCGACTGCTTGGCGCGGATGTCGAAAGCGTGG 743
QY      633 CATCATCTTCAAGGACTGGCTCAAGCTTTACGAAAGATCCGGAACCTCTCTCCCTGTCT 652
Db      744 CTGCTCTGCGCGCACTGGAGCGTGTGCTTACTGCGCGCGCGAGCGCTTGGCGGAGCGCG 803

```

QY 693 GAGCTCTGTACCGGATTTATGTACTCGCGCTGAGCTCTCCGAGAACTGAGCGCA 752  
DB 804 GCGCGTCACTTTCGCGACTA---CTTGCGCACTCTGCGCGCAACGCGCGAGCGCC 860  
QY 753 TCAGCATGATGATTTACTTGAAGCGGCAATCGCGAGCTCCCACTCCGCGAGCT 812  
DB 861 GCGCGGCGCGCGACTTACTGCTGAGAACGCTTGCCTGCGCTTGCAGAGCGCGCGCT 920  
QY 813 TCGATGAAGCGCATCTAATCTGAAGAGATCGCTTCCGACACGAGAGCATG 872  
DB 921 GCGGTGCGCTGCGCGCGAAAGCATCGCGCGCGCGACCGCGCGCTGCAATTCGA 980  
QY 873 GCTGCGCTGAGACTCTGCGGCTGATGTAAGCGGCGCTGCGGAGACCGCGCTGACCC 932  
DB 981 GCTTTCGCGCGGAGAGCGCGCGCTGAGAGCTTTCGCGCGACATGCGCTGACCTT 1040  
QY 933 GACGCGCTCATCTGCGCTGATTTTCCGAGGTGATCGGCGCTGAGCGCGAGCCCG 992  
DB 1041 GTTCAGCGCTGTGCTGCGCTTCCGCGCTTGCCTGCGCGCTGAGAGCGAGCGCGA 1100  
QY 993 GTTTACGCTCAACATTAACGCTTTCACCGGCTCCCGCTCATCGCGCGTGAACGAT 1052  
DB 1101 ATTCTCTCAACGCTCCGCTGTTTCATCGGCAATGCGAGACCGCGATCGCGAGGT 1160  
QY 1053 CACCGGAGCTTACGCTCATGTGCTCTCTGACATCGACACCACTCGCGACAAGCTT 1112  
DB 1161 GATCGCGCACTTACCACTCTGTGTCTGAGAGTCGCGAGTCAGGCGCGGCTGTCTT 1220  
QY 1113 CGAACGCGCGCTTACGCTATTTCAAGAGAGCTGTGAGAGCATGATCACTGCGAGT 1172  
DB 1221 CCGCGAGGCGGTGAAGCTTCCAGCGCACTTCCAGAGCATGACACGCGCATT 1280  
QY 1173 AAGCGTATCGAGTTCAGCGAGGCGCGCGGCTCTGAGGATCCAAAGAGCGCATT 1232  
DB 1281 CCGCGCTGAGAGTCTCCGCGAGCGCGCGG-----CAGGCGCAACGCTCGAG 1335  
QY 1233 GTTCCCGGTGCTCTACGAGCGCGCTTAAACGAGCAATGCTGTGATCACTCTGTGA 1292  
DB 1336 GCGGTGTGTTCGCGAGCAACTGAGCGAGGAGGCTTGTCCGCGCTTTCGCGCA 1395  
QY 1293 GAGGCTCGGAACCTCCGCTGATACACAGACGAGACTCTCTCAGCTCTGTGATCAT 1352  
DB 1396 CGCTTTCGCGCATCTCCAGACATGCTCTGCGAGACCGCGAGGTCTGCTGACACCA 1455  
QY 1353 GCTCTACGAGCAGATGAGGAGCTCTGCTGCTGAGGAGCATGTCAGCGAGTGTCC 1412  
DB 1456 GCTCTACGAGGAGGAGCGGTATCTGCTGCGCTGAGGAGCGGTGCTGTCTCC 1515  
QY 1413 GCGCGCTTTCGAGCAGATGCTCGAGGTACGCTGCTTTCGCGCGCTGACATGA 1472  
DB 1516 CGAAGGCTGCGGAGAAACATGTTCAAGCTACGAGGAGCTCTCAAGGTCTTTCGCA 1575  
QY 1473 GGAACATGAGGAGTGAACAGGTGCTGTTTCGCTTCCGCTGCGCAGCTGAGAGCGG 1532  
DB 1576 CAGGCTGAGGAGCAACCGCGCATCTGCGCTTGCCTGCGCGCAGACAGCGCGCGG 1635  
QY 1533 GAGCGGAGGAGCAGACGCGCTGCTGAGCGAGCATGCTGCGAGCGCTGTTCCGCG 1592  
DB 1636 CTTGCTCAACGAGCAGCGCGCATGCGCACGCGCGCACCTGCTCATGCGCATCTTCT 1695  
QY 1593 GCGGCTGAGCAGCTCCCATGAGCTCGCGGTGCTGCGCGCGGAGAGCGCTGACATGA 1652  
DB 1696 TCGCGCGCGGAGCGCGCGAGTCCGACGCGCTGCTTATCGGACCAACGTCTCACCG 1755  
QY 1653 CGAAGAGCTTTCGCGCTTTCGCGGAGCTTTCGCGCGCGCTGCGGAGCGGAGCGAG 1712  
DB 1756 CCGCGAGCTGCGCGAGGCTGCTGCGCATGCGCGCGCGCTGCGGAGCGCGGAGTGG 1815  
QY 1713 CCGGAAACATTTGCTGCGGTGTGATGAGAAAGCTGAGAGCAGGTGTCTGCGGTCT 1772  
DB 1816 CCTGCGAGCGCGGTGAGGTGACCTGCGCGCGGACCGGACAGAGTCTGCGCGGTATT 1875

QY 1773 CCGGCTGCTGAGTCAAGGCGCGGCTTACGTGCGCATGATGCCGACCTAACCGCGAGCG 1832  
DB 1876 CCGCTGTCTCGCGGAGCGGCTGTACGTGCGCTGAGCATGACAGCGCGCGCAG 1935  
QY 1833 TATCACTACCTCTCTGATCAATGAGTAAAGCTCTGTGACGACGACCATGCTGGA 1892  
DB 1936 GCGGCGCTGATTCGAAGAGGCGCGCGGATGATGCTGCGATTCACGAGAGAGCATTC 1995  
QY 1893 TGGCAACTGTATGCGCGCGGAGATCAAGCGCTGCTGTAAGCAGCGCGCGCTGGA 1952  
DB 1996 GCAGGCTTGTGCGCGCGCTGAGATGACAGCGCTGCGCGCGCGCGCGCTGCGCG 2055  
QY 1993 AGGCGAGCGACACAGCTCCGATGATGCCATTGACACCTTGCATCTCGGTATGT 2012  
DB 2056 CCGGCTGCGCTGCGCGCGCGCAGGCGA-----GTGCTTATGT 2091  
QY 2013 CATCTACACTTCGAGATCAAGAGTTTCCAGAGGAGTATGATCATTCGCGGTG 2072  
DB 2092 GATCTAACACTCGGCTCCACCGGCGTCCAGAGGCGTCAAGGTCAACGCGCGCGC 2151  
QY 2073 CGTCAACCATCTGACATCAACAGCGCTTGAATAGGCGCGGAGACAGGTGCT 2132  
DB 2152 GATCAATACATCGACGCGCTGCTGACCTGCGGCTGGAACGATCGATGCTTCT 2211  
QY 2133 GCGGCTCTCTGCGAGCTTTCATCTCGGTCTATGATGTTTGGGATCTGCGCGC 2192  
DB 2212 GCGGCTCTCGGCTGAGCTTTCATCTCGGTCTTTCATCTGCTTTCGCGCGCTGCGC 2271  
QY 2193 GCGGCTACATCTGCTGCGCGAGCGCTTCAACAGCTGCGCATTCGCGCGCATTTGCGAGA 2252  
DB 2272 CCGTCCAGGCTGCTGCGCGCGCGCGCGCGAGGAGCGCGCATGCTGCTGCGCGG 2331  
QY 2253 GTTATTCAGAGAGAGAGTGAACGCTGTGGAATCTGCTGCGCGCGCTGATGCGATGCT 2312  
DB 2332 GGTATTCAGAGCGGCTGAGGTGAGCTGTGGAATCTGCGCGCGCGCTTGTGGAATGAG 2391  
QY 2313 GGTGAGCATTTTGAAGGCTGCGCGCATTTGCTGCTGCTGCTGCGCGCTTGTGCTGCT 2372  
DB 2392 CTTGAGCTTCCGCGGAGCGAGCGAGCGA-----CATGCGAGCTGCGCGCGCTGCT 2445  
QY 2373 GAGCGCGCATGAGATCCGAGTGGGCTGCTGCGAGCTGCGAGCATCAAGCGC---GG 2429  
DB 2446 GTCCGCGCATGAGGCTGCGCGCTGAGCTGCGCGCGCGCTGCGCGCATGCGCGAGG 2505  
QY 2430 CGTGTGCTGATCACTGCGCGCGGCGCGACCGAGAGCTGATCTGTGCTCATGCGGTACC 2489  
DB 2506 CTGCGCGCTGCAATGCTGCGGTGCGCTACCGAAGCGGAGCATGTGTGAACCTGCGAG 2565  
QY 2490 CGTGAAGAACGTGACCTATCGTGGCGAGCATCCCTACAGCGCGCTGCGCGCAACA 2549  
DB 2566 CTTGATAGGAGTCCGCGCGCGCGCATGCTGCTTCCCTACCGCGCGCATTCGCGAGCA 2625  
QY 2550 GAGCTTCAACGTGCTGATGAGGCGCTGAAACCGCGCGCGCTGCGGTTCGCGGCAACT 2609  
DB 2626 GGCCTACGCGGTGTGCAACCCAGCGCGCGAGCTGCGCGCGCTGCGGTGCGGAGCT 2685  
QY 2610 CTACATGCGCGGCTGCGCGCGCGCATGCGGCTGCTGCGGAGTGAAGAGAGCGGAA 2669  
DB 2686 GTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCATGCAACGATTCGCACTCAGCGCG 2745  
QY 2670 GAGCTTCTCTGCGAGACCGCGAGACCGGAGCGCTTCAAGACCGCGCATCTTGGCG 2729  
DB 2746 GCGCTTCTG-----CAAGATGCCAGGCGCGCTGATTCGCAACCGCGCATTCGCGGT 2799  
QY 2730 CTACCTGCGCGATGAAACATGAGTTCATGAGGAGTGAAGCAACCAATCAAGCTTTCG 2789  
DB 2800 CTACTGCGCGCGAGGATCCCTGAAATCTCGGTGCGGTGCGACAGAGGTGAAGTGG 2859  
QY 2790 CGGATTCGCGGTGAGCTGCGGAGAAATGAGAGAACTCAAGTGCATCCGAAGT 2846  
DB 2860 CCGCGAGCGCATGAGTGTGGGCGAGGTGAGAGCGCGCGCTGCGCGCGCGCTGCGGT 2916

RESULT 12  
US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6294338  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 8.5%; Score 358.2, DB 3; Length 4411529;  
Best Local Similarity 48.4%; Pred. No. 7.2e-69;  
Matches 1252; Conservative 0; Mismatches 1283; Indels 54; Gaps 7;

```
QY 286 CACGCTTACGGGATAGACATGTAACGATCTCGACGTGCGAGGCTGAGCGCGCCCTTT 345
DB 2675453 CACCTTACGTCGATTCATGATGTCGGGATGTCGAGACAGGCTCGCGCGCGCC
QY 346 CGGAAAGTCGTGCGCGGACGACATGCTTCGGGCGCACACGCTCGCGACATGATGAG 405
DB 2675403 ACCCGATTGGGCGCTCGGACCGGATGCTGCGGGGAGATTCTTGCGGACCGACCG 2675344
QY 406 GTGATCGAGCTTAAAGTCGACCGCGACATCGAGATCATGATCTGCGGGGCTCGACCG 465
DB 2675343 CGCATTCGCGCGCGCGGAGATCTGCGACCTTCCATCATGATGTTGCCACCTGCTC 2675284
QY 466 AGCACAAGGGAAGGAGGCTGCTGTCGAGATGCGATTCGACCGCATCTATATAC 525
DB 2675283 GTGGCCCGGATGTCGTCGATGACGAGGTTGCGGGAGATTCGCGACCGCAATGCA 2675224
QY 526 ACCGAGCGCCCTCGCTCTATCAGCTGCTGCGCTTGGCTGAGACGAGCGGAAACCGT 585
DB 2675223 CAGCTCGAGCGGTGGGATTCGAACTTGCTGATGATTAACCGGGGAGCGCACCGC 2675164
QY 586 CTGCTGCTCAGTATGATCTCAATTAAGTTGACCTAGGAGCGCTGCTCATCTTCAAG 645
DB 2675163 CTACATGTCGACCTGGAATGAGGCGCGTGAACGAGATGACTACCGCATCTTGCT 2675104
QY 646 GACTGCGTCAAGCTTCTACGAAATATCCGAGACCTCTCTCCCTGCTGAGCTCTCGTAC 705
DB 2675103 GACTGCGCGGCTCTTATGACGCGGTGAACCGCGGCACTGGGGCTTACACTTAC 2675044
QY 706 CGCGATTATGTAATCTGCGCGTGAATCTGCAAGAGTCTGAGCGCATCAAGATGATG 765
DB 2675043 TACCGGAGGCTATGAGGCGGAGAGACGCTGCCCAACCGGTTTCGAGCGCGAC 2674984
QY 766 GATTACTGGAAGCGGCGCATGCGGAGCTCCCACTCGCGGAGCGCTTTCGATGAAGGC 825
DB 2674983 GACTGATGAGCGGACGATCCCGGAGCTGCGGATTCGCGCGCTGCTACCGGGCC 2674924
QY 826 GATCATCATACCTGAAAGAGATTCGGCTTCGGGCAACGAGACATGATGCTCCGTCGAC 885
DB 2674923 GCGCGGAAACGAGACCGGGCGCGGACACCGCGCC--TGGCACTGGGTAGAACCCGAG 2674867
QY 886 TCTTGGGATGATTAAGCGGCTGTCGGGAGCGCGGCTGATACCCGACGCGGCTCATC 945
DB 2674866 ACCCGGACGCGGTTTGGCCGAGCCCGGGGATGATACCCCGCGGATGACGCTG 2674807
QY 946 CTGGCTGATTTTCCGAGGTGATCGAGGCTGGAAGCGCGAGGCCCGCGTTTACGCTCAC 1005
DB 2673744
```

```
DB 2674806 GCGCGGCTTTGCCAAGTGTGCGGCTGCTGAGGCTGTCAGGCTGTCGCGGCTCTCTGTAAC 2674747
QY 1006 ATTAAGCTTTCAACCGGCTCCCGTCCATCCGCGGAGTAAGATATCAACCGGAGACTTC 1065
DB 2674746 CTGCGGTTGTTCACTGCGGACGAGCCCTGATCAACAGCTGCTGCTGCTGATCTTC 2674687
QY 1066 ACGTGATGTCCTCTGAGATGACACCACTGCGGACAAAGAGTTGAAACGCGCT 1125
DB 2674686 ACTCTCGCTGTTGCTGATGATGATGACGAGTGGCGGACCGGCGCGGCGGCGG 2674627
QY 1126 AAGGTATTAAGAGACGCTGAGGAGGATGATGATCTGAGACTGAAGGCTATCGAG 1185
DB 2674626 CAGGCGGACGAGAAACCTTGAGAGCGCCCGAGGCGCAAGTGCATACCCGCGCTCT 2674567
QY 1186 GTCCAGGAGAGGCGCGCGGCTCTGAGGAGATTCGAACGAGGCGCATTTGCTCCCGTGG 1245
DB 2674566 GTGCTGCTGATCTGAGCGGCGGACCGGTGACCCAGGTGCTGACCGGCTGATTCAC 2674507
QY 1246 CTCAAGAGCGGCTTAAACAGCAAGTCGTTGCTGCTGCTGCTGAGAGCTCGGAAT 1305
DB 2674506 AGCGCGTGGGCTCGCGGACCTTTTCTGCGCGGACGTCAACGAGCA--ATTGGGACCA 2674450
QY 1306 CCGGTGACACGAGACGCAAGCTCTGAGCTGCTGCTGATCATCACTCTTACGAGAC 1365
DB 2674449 CCGGATGATCATTTTCCAGAGGCGCCGAGGTGCTGCTGACGCGGAGTCAACGAGTT 2674390
QY 1366 GATGGGACCTGCTGCTGCGGAGGACATGCTGACGAGAGTGTTCGCGCGGACCTTCTG 1425
DB 2674389 GACGCGGTGCTGCTGATGATGATGATGCTGCGGAGGAGGCTTTTGACCCGCGCT 2674330
QY 1426 GACGATGCTGGAAGCTGACGTCGTTTCTCGGCGGCT-----CACTGAGGAACA 1479
DB 2674329 GACGCTATGTTTACCAACAGAGTGAAGATTTGCTCGGTTGGCGCGGAGGACAGCGC 2674270
QY 1480 TGGGATGACAGGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1539
DB 2674269 TGGGATGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2674210
QY 1540 AACGCGACCAAGCGGCTGCTGAGGAGATAGGCTGACGCGGCTGCTGCGGCGGCGG 1599
DB 2674209 AACGTCGACCGCGCGCGCGCGGACGAGCGGCTGACGAGCGGCTTTTCCGCGAG 2674150
QY 1600 GAGGAGCTGCCATGACGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1659
DB 2674149 CAACAGAGCGCGGACGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2674090
QY 1660 CTTCGCGCGCTGCGGCGGACTTGGCGCGCGCTGCGCGGCGGACGCGCGGAC 1719
DB 2674089 CTGCGGACGAGGATCGCGGCTGCGCGCGCTGCTGCTGCGGCGCTACGAGTCGC 2674030
QY 1720 ACATTGCTGCGGCTGATGATGAGAAAGCTGAGGAGCGGCTGCTGCGGCTTCTGCG 1779
DB 2674029 GACACGTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673970
QY 1780 CTGAGTCAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1839
DB 2673969 TTGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673910
QY 1840 TACTCTCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1899
DB 2673909 CGCATCTCTGAGGACCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673865
QY 1900 CTGATGAGCGCGGAGTCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1959
DB 2673864 CCGCATGCAAGTGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673805
QY 1960 GCGACAGCGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2019
DB 2673804 GCGCGGAGAAATTCGTCGCGGCTGAGGATTCACCGCGCTGCTGCTGCTGCTGCTG 2673745
QY 2020 ACTTCGAGATCAACAGGCTTCCCAAGGAGGATGATGATGATGATGATGATGATGATG 2079
DB 2673744 ACTTCGAGCTCAACCGGAGAACCAAGGATGATGATGATGATGATGATGATGATGATG 2673685
```



Db 2671446 CCAACGACGACCCGACGCGCGCGGCTGTTGCGCAGTTCCGGCGACTTGAAGCTACGCCCA 2671387  
QY GCTTTGCGCGGCTTGGCGCGCGGCACTTGGCGCGCGGCTGCGCGAGAGAGGCGACGCGCGCA 1718  
Db 2671386 ACTGGCGACCAAGGATGCGCGGTGCGCGCGCGCTGGCTGCTGGCGGCTTACAGAGTGG 2671327  
QY 1719 CACATTGGTGGCGGTGATGAGAGAAAGCTGGAGAGAGTTGTCGCGGTTCTCGCGGT 1778  
Db 2671326 CGACACCGTGGCGGTGCTGGGTCGAAAAAGGCGAACAAGTGGCGGCTGTGCTGGGAT 2671267  
QY 1779 GCTGAGTCAGCGCGGCTTACGTCGATCGATCGATCGATCGATCGATCGATCGATCGA 1838  
Db 2671266 TTTGGCGCGCGGCTTACGTCGATCGATCGATCGATCGATCGATCGATCGATCGA 2671207  
QY 1839 CTACCTCTCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGA 1898  
Db 2671206 GCGATCTCTGGGACGCGGCTTGGCTTACTT-----AGGCTCTTTGGCG 2671162  
QY 1899 ACTGTCATGGCGCGCGGAGATCGAGCGGCTGCTGTCGAGCGAGCGCGCGCTGAGAGCGA 1958  
Db 2671161 CCGCGCATGCAAGTGGCGGCTGCGGTCGCGACGCTGTTGCTGCTGAGAGTCTTGGCGC 2671102  
QY 1959 CGGCGACCAAGCTCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2018  
Db 2671101 CGCGCGCGAGATTCGTCGCGCGGCTGAGCGATCCACCGCGCTCGCTTATGTTGTT 2671042  
QY 2019 CACCTCGGAGTTCACAGGAGTTGCGCGAGGAGTGTATGATGATGATGATGATGATGATGAT 2078  
Db 2671041 CACCTCGGAGTTCACAGGAGTTGCGCGAGGAGTGTATGATGATGATGATGATGATGATGAT 2670982  
QY 2079 CACCATCTGAGATCAAGAGCGCTTGGAAATAGAGGCGCGAGAGAGAGAGAGAGAGAGAG 2138  
Db 2670981 CACCGTGAAGACTTATCTCCGCACTTCCGAGCTAGAGCGCGAGATGCGGCTTGGCTT 2670922  
QY 2139 CTCCTCGTGAAGTTCATCTCTCGCTATGATGATGATGATGATGATGATGATGATGATGAT 2198  
Db 2670921 GCGCAGCTGAGAGTGGCAATGTCGCTGTCGATCTTCCGCGCTCGCTCGCGCG 2670862  
QY 2199 TACGATCTGATGTCGCGAGCGCTTCCAGCTGCGAGCTCGCGAGTATGATGATGATGATGAT 2258  
Db 2670861 AGCGATCTGATGTCGCGAGCGAGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 2670802  
QY 2259 CGAAGAGAGAGAGTGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2318  
Db 2670801 CGACATTAAGAGTGAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2670742  
QY 2319 GCATTTTGAAGTGTGCGCGGATTCGCTGATGATGATGATGATGATGATGATGATGATGAT 2378  
Db 2670741 AGTCGCGGAGGCGCGCTG-----TCGTCGCTGCGAGCGAGTGTGCGCGCG 2670694  
QY 2379 CGACTGATTCGCGGAGTGTGCGCTGTCGAGCTTCCAGGCAATCAGGCGCGGAGTGTGCGT 2438  
Db 2670693 CGACTGATTCGCGGAGTGTGCGCTGTCGAGCTTCCAGGCAATCAGGCGCGGAGTGTGCGT 2670634  
QY 2439 GATCAGCTGCGGAGGCGCGAGCGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2498  
Db 2670633 CGCGGAGTGTGAGGCGCGAGCGGCTTCCAGGCAATCAGGCGCGGAGTGTGCGT 2670574  
QY 2499 CGTGCAC-----CTATGTCGCGAGAGTTCCTTACGCGCGCTCGCTGCGAGCA 2549  
Db 2670573 CGCGCGCAATGTGCGACCGAGCTGCGTGTGCGATCAGGCGCTTCCGTTTCCCAACA 2670514  
QY 2550 GACGTTCAAGTGTGATGAGGCGCTGGAACCGGCGCGGCTGCGGAGTGTGCGGAGCACT 2609  
Db 2670513 CGCTGCGGAGTGTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2670454  
QY 2610 CTACATGCGCGGAGTGTGCGGCTGCGAGCTGCGGCTTACGCGAGTGTGCGGAGCA 2669  
Db 2670453 GTGGGTGTCGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAG 2670394  
QY 2670 GAGCTTCTCTGTCACCGCGAGAGCGGAGCGCTTACGAGAGCGGCGATCTGCGCG 2729  
Db 2670393 GCGCTTCTCTGTCAGTGAAGCGGCGGAC-----CTGTATGCGCACCGGCTGATGTCGCGCG 2670340

QY 2720 CTACCTGCGCGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCG 2789  
Db 2670339 CTACCTGCGCGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCG 2670280  
QY 2790 CGGATCGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCG 2849  
Db 2670279 CGGATCGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCG 2670220  
QY 2850 CGAGCGCGGTG 2859  
Db 2670219 CGCGCGGCG 2670210

RESULT 14  
US-09-252-991A-4678/C  
; Sequence 4678, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107136.136  
; CURRENT FILING DATE: US/09/252, 991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4678  
; LENGTH: 1842  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4678

Query Match 8.2%; Score 348.6; DB 4; Length 1842;  
Best Local Similarity 51.7%; Pred. No. 1.6e-67;  
Matches 946; Conservative 0; Mismatches 839; Indels 44; Gaps 5;

QY 1064 TCAGTGTGATGTCCTCTGAGATCGACACCACTCGGACAGAGCTTGGAGACGCGG 1123  
Db 1842 TCACACCTGTCCTGTCGAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAG 1783  
QY 1124 CTAGCGATTCAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 1183  
Db 1782 TGAAGAGTTCACGCGCACTTCAAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 1723  
QY 1184 AGTTCAGAGAGAGCGCGCGGCTGTCGAGATTCAGAGAGCGCATTTTCCCGTGG 1243  
Db 1722 AGTGTCTCGGAGAGCGCGCGG-----CAGGGCGAGCGACGCTCGCGCGGAGTGT 1668  
QY 1244 TGTTCAGAGAGCGCGCTTACAGCAAGTGTGTCACCTCGTTGAGAGGCTCGGA 1303  
Db 1667 TCGCAGAGACCTGCGAGAGAGAGGCTTGTTCGCGCGCTTCCGACCTTTCGCG 1608  
QY 1304 CTCGAGTGTACAGAGAGCGAGACTCTGAGTGTGTCGATGATGATGATGATGATGATGATGATGATGAT 1363  
Db 1607 ATCTTCAGAGATCTCTGCGAGAGCGCGAGTGTGCTGAGACCAAGCTTACCGGG 1548  
QY 1364 AGATGAGAGAGCTGTCCTCGTGTGAGAGATGTCGAGAGTGTTCGCGCGAGCTTC 1423  
Db 1547 TGGGCGAGAGTATCTGTCGCTGCGAGAGCTGTGTCGCTGTTCGCGAGAGCTTC 1488  
QY 1424 TGAAGAGATGTCGAGAGAGTGTGTCGCTGTCGCGCGGCTTCACTGAGAGAGCTTCG 1483  
Db 1487 CGGAGAGATGTCGAGAGAGTGTGTCGCTGTCGCGCGGCTTCACTGAGAGAGCTTCG 1428  
QY 1484 GTGAAGAGTGTGTCGCTGTCGCTGTCGCGCGGCTTCACTGAGAGAGCTTCG 1543  
Db 1427 AGCAGCGCGAGTGTGCTGTCGCTGTCGCGCGGCTTCACTGAGAGAGCTTCG 1368  
QY 1544 GACCAAGCGCTGTGAGAGAGATGCTGCAAGCGCTGTTCGCGCGGCTGAGC 1603



Db 1367 GCCAGCCGGCAGTGCGCCCAACGGGCGGCCAACCTTGCAGACCTTCTTCTTGGCGCGCG 1308

QY 1604 AGCTGCCATGCACTCGCCCGTGTGTCCGGCGCGCAGACGCTTACGTACAAAGACTTT 1653

Db 1307 AGGCGCCCGGATGCCAGACCGCGCTGTCTTATCCGACCAACGTGTACACCGCGCGCAACTGG 1248

QY 1664 CGCGCGCTTCCCGGGAATTGGCCGGCGCGCTCCGCGAACCAGGGGCAAGCCCGAACAT 1723

Db 1247 CCGAGCGGTCCGTGGCAATCCGCGCGCGCCTCGCGGAAGCCGGGGTGGCCCTTGGCGAG 1188

QY 1724 TGGTCGCGGTGTGATGAGAAAGGCTGGAGACAGTGTGTGCGCGTTCTCGCGGTCTCG 1783

Db 1187 CGGTGAGGTCAAGCTTGCCTGCGCGCGGACCGCAGCAAGTTCGGCGGTATTTCGGCGTCTCG 1128

QY 1784 AGTCAGCGCGGAGCTTACGTGCCGATCGATGCCGACTTACCGGCGGAGGTATTCATACC 1843

Db 1127 CCGCAGGCGCGCTGTACTACGTGCGCTGTGACATTCGACCAACCGCGCGCACGGCGCGCTGA 1068

QY 1844 TCCCTGATCATGTGTAGGTAAAGCTCTGTGTGACGCAACCCATGGTGTGATGGCAACTGT 1903

Db 1067 TCGAAGAGGCGCGCGGGTATGTCTCGGATCATCGAAGAGAGCATCCGAGGCTTTC 1008

QY 1904 CATGGCCCGCGGGAGTTCAGACGCGTGTCTGTAGAGGAGCGCGCGCTCGAAGCGCACGCG 1963

Db 1007 CGCGCGCGCTGTGATTTCCAGCGCCTGTGTGGCGCGCGCGCTGACCGCCCCG----- 954

QY 1964 ACCAGCTTCGATGTATGCCATTACAGACCTTCGGATCTCGCGTATGTATCTTACACTT 2023

Db 953 -----TGCCTGTGGCGCGCGAGGCGAGTGCCTATGTATCTTACACTT 912

QY 2024 CGGATTCACAGGCTTGGCCCAAGGGGTGTGATGTGATCATTCGGGTTCCGTCAACACA 2083

Db 911 CGGGTTCACCGGGTGTGCCAAAGGCGGTGTGAGTCAAGCACAGCGCGCGCATCAATACA 852

QY 2084 TCTTGGACATCAACAGAGGCTTCGAAATATAGAGCGCGGAGACAGGGTGTGGCGCTCTCT 2143

Db 851 TCGAGCGCGCTGTGCACTGCTGCGGGTGTAAAGCATCGGATTCGCTTGGCGGTCTTCG 792

QY 2144 CGGTAGCTTCGATCTCTCGGTCTATATGATGTGTTCGGGATCTTCGCGCGCGGCGGTACA 2203

Db 791 CGCTGACCTTCGATCTGTCTGTGCTTCGACTGTTCGAGCGCGGTGCTGTCGCGGCTCAAC 732

QY 2204 TCGTGTGTCCCGAGCGCGTTCCAAGCTTCGCGGATTCGGCGCATTTGGCGAAGTTATCGAAC 2263

Db 731 TGTGTCCTCCCGGCCAGGAAACAGGCGCGCGATGCGCGTGTGCGCGGCGGAGGCTATCCAGC 672

QY 2264 GAGAGAAAGTGCAGGTGTGTGGAATCGGTGTGCGCGCGCTGATGCGATGTGCTGTGCGACTT 2323

Db 612 GGCATTCGGGTAGGCTGTGTGAATCTGCGCGCGGCTTGTCTGAGATGTGCGCTTCAGCTTGC 612

QY 2324 TTGAGGATCGCCCGCATTCGCTCGCTAGATCTTCGCGGCTTTCGCTGTGAGCGCGCAT 2383

Db 611 CGGCGAGCGAGCCGA-----CTATGCAAGTCTGCGGCGGTTGCTGTGTCGCGCAT 558

QY 2384 GGATCCGCGGTGGGCTTGCCTGGCGGAGCTTCAGGCGCATAGGCGC---GGCGTGTGTA 2440

Db 557 GGGTGGCCTCTGGACTTGCCTGCGCGCGCTGTGCGCCACTGTGTGTGCCAAGGCTTGCCTGTGC 498

QY 2441 TCAGCTTGGGGGGGCGCAACGAGGCTGTGATTCGATCGGAGTACCCCGTGAAGAGAG 2500

Db 497 ATGTGCTGGGTGGCCCTAACCGAAGGGGCAATGTGATCGAACTGTGAGAGCGTTCGATACG 438

QY 2501 TCGACTTATCTGTGGCGGAGCAATCCCTTACCGGCGCTTCGCGTGGCAACAGACGTTCACG 2566

Db 437 TGCGCCCGCACTGGGCTTCGATTTCCCTTACGCGCGGCAATTCGCGGGAACAGGCTTACCG 378

QY 2561 TGTCTGATGAGGCGCTCGAACCGCGCGCGGTCTGGGTTCGCGGCAACTTACATTGGCG 2620

Db 377 TGTGTGAACCAACCGAGGCGCGACGTGCGGAGCTGTGTGTGGCGAGCTGTGTGATTCGCG 318

QY 2621 GGGTGGGCTGGCACTGGGCTTACTGGCGCGATGAAAGAGACGCGCAAGGCTTCTTCG 2680

Db 317 GCGCAGACTGAGCCCGGCGTATGCAACGATCCCGAACTCAGCGCCCGCGCTTCGT-- 260

Qy 2681 TGCACCCCGAGACCGGGGACGCGCTTACAAAGACCGGCGATCTGAGCGCGCTTACCGCCG 2740

Db 259 ---CCACAGATGCCCAAGGCGCGCGCTGGATGCGACCGGCGAATCGCGTGTCTACTCGGGCG 204

Qy 2741 ATGGAACATCGATGATTGATGGGCGGTAGGACAACCAATCAAGCTTTCGGGATCCGCG 2800

Db 203 ACGGTATCCGTGGAATTCCTCGGTGGGGTCGACCGACGAGTGGAAAGTGCGGCGGCACGGCA 144

Qy 2801 TTGAGCTCGGGAAATCGAGGAACGCTCAAGTCGATCCGAACGTACGGGACGGCGTGA 2860

Db 143 TCGAGTTGGGGAGAGTAGAGGCGCGCGCTGGCGCCAGGCTGGCGTGATAGCGCCTTCG 84

Qy 2861 TTGAGCCGTCGGGAAACGACGCGCGAAC 2889

Db 83 CGGCGGTGCTCGGCGGTGCGCTGGCAGC 55

```

RESULT 15
US-09-252-991A-4764
; Sequence 4764, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4764
; LENGTH: 2844
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4764

Query Match          8.0%; Score 340; DB 4; Length 2844;
Best Local Similarity 51.9%; Pred. No. 1,3e-65;
Matches 9177, Conservative 0; Mismatches 805; Indels 44; Gaps 5;

QY      1127 AGCCTATTCAGAGACACCTGTGGAGACGATGATCATCTGCGACGTAAAGCGTATCGAGG 1186
DB      . 1 AGACTTTCAGCGCAACTCCAGGACCATGACACGCGCGCATTTCCCGCCTGGAGG 60

QY      1187 TCCAGCGAGAGGCGCCCGGGTCTGTGGGATCCACAGAGCGCATTTTCCCGTGTGC 1246
DB      61 TGCTCCGAGGCGCGCGG---CAGGGGCCACGCCACCTCGGCGCGGTGGTTCG 115

QY      1247 TCACGAGCGGCTTAACGACGAAGTCGTGTGATCACTCGTTGACAGAGCTGCGAATCTC 1306
DB      116 CCACGAACCTGGGAGAGAGGGCTTCCTCCCGGCGCTTCCGAGCGCTTTGGGCATC 175

QY      1307 CGGTGTACACCAAGCAGCAGACTCTCAGCTGTCTGTGATCATCACTCTACGACAGC 1366
DB      176 TCACGACATGCTCTCGAGAGCCCGGAGGTCTGGCTGACACCAAGCTCTACCGGGTGG 235

QY      1367 ATGGGAACTGTCTCTGCGCGTGGGACATGCTCAACGAGAGTGTTCGCCCGCACTTTTGG 1426
DB      236 GCGACGGTATCTCTGCTGGCGCTGGGACAGCGTCTGCGCTGTTCGCCGAAGGCGTCGCG 295

QY      1427 ACGACATGCTCGAAGCGTACGTTGTTTTCTCCGGGCGGCTCAGTCGAGAAACATGGGGTG 1486
DB      296 AAACCATGTTGGAAGCTACGATGGGCTGTCTCCACGCTCTCTGCGACAGCGCTTGGAGC 355

QY      1487 AACAGTGTGCGTGTCTTCCTTCGCGCTCCAGCTAGAAAGCGCGGCGAGCGCAACCGCA 1546
DB      356 AGCCGCGCATCTCCCTTGTGCTTGGGCGGACGAGCGGCGCGCGGCGCTGTCAACGCGC 415

```

QY 1547 CCAACCGCTGTGAGCGACATACGCTGACAGCGCTTTGCGGGCGGGGTGAGCAGC 1606  
Db 416 AGCCGCGATGCGCCACGCGCGCGCACCTTGCAATCGGACCTTCTTCTTGGCGCGCGCGAG 475  
QY 1607 TGCCCATGACGCTGCGCGGTGTGCGGCGCGCAACCTTCACTGACGAAAGCTTTGGC 1666  
Db 476 CGCCGAGATGCGCACCGCTGTCTATCCGACCAAGTGTCAACCGCGCGCAACTGCGCG 535  
QY 1667 GCGCTTCCGCGGACATTTGCGCGCGCGCTGCGGACAGAGGCGCGCGCGCAACATTTG 1726  
Db 536 AGCGTGGCTGTGCAATCGCGCGCGCGCTGCGGAAAGCGGGGTGGCGCTTGGCGACGCGG 595  
QY 1727 TCGCGGTGTGATGAGAAAGCTGGAAGCAGAGTTGTGCGGTTCTCGCGGTCTGAGT 1786  
Db 596 TCGAGGTGAGCTGCGCGCGCGCGACCGCAGAGTGTGCGGCGGTATTCGCGGTCTGCGCG 655  
QY 1787 CAGGCGCGCGCTGACGCGCGATTCGATGCGCGACCTTACCGCGGAGGCTTACCTACCTCC 1846  
Db 656 CAGGCGCTGTCTACGCGCGCTGAGCAGTACGACCAAGCGCGCGCGCGCGCGCGCTGATCG 715  
QY 1847 TCGATCATGTGATGAGGTAAAGCTGATGAGCGAGCGACCAATGCTGATGAGCAACTGTAT 1906  
Db 716 AAGAGGCGCGCGGGGTATGCTGCGGATCACCGAGAGAGACGATCCGAGGCTTGGCGCG 775  
QY 1907 GAGCGCGCGGATCCAGCGCGCTCTGTGAGCGAGCGCGCGCTCGAAGCGCGAGCGCAAC 1966  
Db 776 CGCGCTGAGATGTCCAGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826  
QY 1967 AGCTCCGATGATGCGCCATTGAGACCTTGGGATCTGCGGTATGTCACTTACACTCGG 2026  
Db 827 -----TGCCTGCGCGCGCGCGAGCGAGGTGCTATGTATCACTACCTCGG 871  
QY 2027 GATCCACAGGGTTGCGCGCGAGGGGTGATGATGATCATCGGGGTGCGCGCAACCATCC 2086  
Db 872 GCTCCACCGGGGTGCGCGCGAGGGGTGAGGTGAGCGACGCGCGCGCGATCAATACATCG 931  
QY 2087 TGAACATCAACGAGCGCTTCAAAATAGGGCGCGAGACAGAGGTGTGCGCTTCTCTCGC 2146  
Db 932 ACGCGCTGTGACCTGTGCGCGGTGAACGCAATCGGATCGCTTGTGCGCGGTCTCGCGC 991  
QY 2147 TGAAGTTGATCTCTCGGTCTATGATGTGTTCGGGATCTTGTGCGCGCGCGGTACGATCG 2206  
Db 992 TGAATTGATCTCTCGGTCTTTCGACTGTTCGCGCGCTTGTGCGCGCGCGGTGCGCGCGCTCG 1051  
QY 2207 TGTGCGCGGACGCGTCCAGCTGCGCGATTCGCGCGCATTTGGCGAGTTGATCGAACGAG 2266  
Db 1052 TCTGCGCGCGCGAGAACAGGCGCGGATGCGCTGCTGCGCGAGGCTATCCAGCGCGC 1111  
QY 2267 AGAAGGTACGAGTGTGAATCTGAGTCCGCGCTGATGCGGATGCTGTGAGCATTTTG 2326  
Db 1112 ATGCGGTAGCTGTGAAATCTGGCGCGCGCTTGTGGAAGTGGCCTCAGCGCTGCGCG 1171  
QY 2327 AGGTTGCGCGCGATTCGCTCGGTAGGTCTGTGCGCGCTTCTGTGAGCGCGCATGGA 2386  
Db 1172 CGAGCGAGCGCGA-----CTATCGAGTGTGCGCGCGGTGTGCTGTCTGCGCGCATGCG 1225  
QY 2387 TCCCGGTGCGCGCTGCTGCGCGAGCTCAAGCGCATCAGGCGC---GGCGGTGCGGTATCA 2443  
Db 1226 TGGCGCTGGAAGCTTGCCTGCGCGCGCTGCGCGCATTTGTGCGAGAGCTGCGCGCTGCAATG 1285  
QY 2444 GCTTGGCGCGGAGCCACCGAAGGTGATGTGTTCATCGGATACCGGTGAGGAAGTTCG 2503  
Db 1286 TGTGCGGTGTGCGCTACCGAAGGGGCACTGTGTCAACCTGCAAGCGTCAATACGTGC 1345  
QY 2504 ACCTATGCTGCGCGAGCATTCCTTACGGCGCTTGTGCGCGCAACCAAGCTTCAAGTTC 2563  
Db 1346 CGCGCGACTGTGCTGATTCCTTACGCGCGCGCATTTGCGGAGCAGGCGCTTACCGGCTG 1405  
QY 2564 TCGATGAGGCGCTCGAACCGCGCGCTGTGCGGTTCGCGGCACTTACATTTGCGGGG 2623  
Db 1406 TCGACACCCAGCGGCGCGAGCTGCGCGAGCTGTGCGGAGCTTGTGATCGCGCGCG 1465  
QY 2624 TCGGCGGTGCACTGTGGCTACTGTGCGCGATGAAGAGAGCGCGCAAGCTTCTCTGTGC 2683

Db 1466 CCAAGCTGCGCGCGCGCTATGCAACGATCCCGAACTCAGCGCGCGCGCGCTTCTG----- 1520  
QY 2684 ACCCGAGACCGGGGAGGCGCTTACAAAGACCGGCGATCTGAGCGCTTACCTGCGCGATG 2743  
Db 1521 -CCAGATGCGCGCGCGCGCTGTGTGCGACCGGCGATCGCGGTGCTTACGCGCGAGC 1579  
QY 2744 GAAACATGAGTTGATGCGGCGGTGAGGACCAATCAAGCTTTCGCGGATACCGCGTTG 2803  
Db 1580 GTACCTGGAATTCCTGCGGTGCGGTGCAAGCAGAGGTGAAGTGTGCGCGCGCGCATCG 1639  
QY 2804 AGCTCGGGGAAATGAGGAAAGCTCAAGTGTGATCCGAAGTACCGACCGCGGTATG 2863  
Db 1640 AGTTGGCGAGGTGAGGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1699  
QY 2864 TGCCGCTGCGGAGACGCGCGCGCAAC 2889  
Db 1700 CGGTGCTGCGCGGTGCGGTGCGCGAGC 1725

Search completed: October 4, 2003, 03:40:15  
Job time : 267.251 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: October 3, 2003, 11:29:26 / Search time 1035.8 Seconds  
(without alignment)  
11031.749 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_11872\_16104

Perfect score: 4233  
Sequence: 1 atgcagcatcaccgctctc.....agggcagagacgacgactaa 4233

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4233	100.0	68750	21	AAZ55887
2	4114.6	97.2	71989	21	AAA29349
3	728	17.2	58857	21	AAA58471
4	543.4	12.8	4305	22	ABA89198
5	543.4	12.8	48715	22	ABA89188
6	543.4	12.8	50538	24	AB878890
7	390.4	9.2	32160	24	AB878887
8	360.6	8.5	135638	25	ABX34289

C	9	358.2	8.5	4411529	22	AA199682
C	10	349.2	8.2	4403765	22	AA199683
C	11	305.6	7.2	1416	23	AA53964
C	12	284.2	6.7	7335	22	AA81367
C	13	265	6.3	37856	21	AAA11992
C	14	261.4	6.2	13029	23	AA51470
C	15	247.6	5.8	30600	24	ABQ78872
C	16	245	5.8	7110	22	AA81361
C	17	241.6	5.7	77536	21	AAA14651
C	18	239	5.6	88421	24	AA140781
C	19	221.8	5.2	7347	22	AA54136
C	20	220.6	5.2	5305	22	AB89272
C	21	211.2	5.0	1566	22	AA826321
C	22	203.8	4.8	135638	25	ABX34289
C	23	202.4	4.8	5451	22	AA826319
C	24	202.2	4.8	1706	22	ABA89271
C	25	188	4.4	3348	22	ABA89272
C	26	181.4	4.3	1341	22	AA826322
C	27	179.4	4.2	18660	21	AAA58472
C	28	177.4	4.2	4341	25	ABT32151
C	29	177.4	4.2	32539	25	ABT32129
C	30	172.6	4.1	7178	21	AAA59145
C	31	172.2	4.1	3849	22	AA825795
C	32	171.6	4.1	23666	12	AAQ10190
C	33	168.4	4.0	8301	22	AA826035
C	34	168.4	4.0	34071	22	AA826033
C	35	168.4	4.0	42717	22	AA826032
C	36	158.2	3.7	1178	20	AA826032
C	37	155.2	3.7	3471	24	ABX74876
C	38	155	3.7	1200	20	AA826035
C	39	152.6	3.6	6390	22	AA81370
C	40	150.8	3.6	1204	20	AA826035
C	41	148.4	3.5	1803	25	ABT32150
C	42	148.6	3.3	31122	14	AAQ40706
C	43	137.4	3.2	1172	20	AA826032
C	44	137.4	3.2	4615	24	AA826034
C	45	136.6	3.2	3798	22	ABX74884

#### ALIGNMENTS

AAZ55887	standard; DNA; 68750 BP.
AAZ55887	
10-APR-2000	(first entry)
Sorangium cellulosum 68.75 kb contig.	
Epithelone biosynthesis; type I polyketide synthase; taxol substructure; anticancer; ds.	
Sorangium cellulosum.	
Location/Qualifiers	
1..1826	
/*tag= a	
/product= "partial Orf 1 protein (AAZ58580)"	
/note= "No initiation codon given in the specification"	
complement (1900..3171)	
/*tag= b	
/product= "Orf 2 protein (AAZ58581)"	
3415..5556	
/*tag= c	
/product= "Orf 3 protein (AAZ58582)"	
complement (5612..5992)	
/*tag= d	
/product= "Orf 4 protein (AAZ58583)"	
6226..6675	

Mycobacterium tube  
Mycobacterium tube  
Klebsiella pneumoniae  
Quorum sensing con  
S. cellulosum DNA  
Pseudomonas aeruginosa  
S. roseosporus dep  
Quorum sensing con  
Nucleotide sequenc  
88421nt genomic DN  
Pseudomonas aeruginosa  
Escherichia coli p  
Pseudomonas sp 11p  
Pseudomonas sp 11p  
S. atroolivaceus 1  
Pseudomonas sp 11p  
Escherichia coli p  
Escherichia coli p  
Pseudomonas sp 11p  
Nucleotide sequenc  
Benzodiazepines b1  
Benzodiazepines b1  
DNA encoding a pep  
S. chrysomallus ac  
Cephalosporin anti  
Nucleotide sequenc  
Nucleotide sequenc  
Soil derived pep1  
Soil derived pep1  
Bacillus lichenif  
Soil derived pep1  
Quorum sensing con  
Soil derived pep1  
Benzodiazepines b1  
Bacillus subtilis  
Soil bacteria pept  
Nucleotide sequenc  
Bacillus lichenif

```

FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
CDS
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /tag= v
FT      /product= "Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX
XX      WO9366028-A2.
XX
XX      23-DEC-1999.
XX
XX      16-JUN-1999; 99WO-EP04171.
XX
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
XX      05-FEB-1999; 99US-0118906.

```

```

PA      (NOVS ) NOVARTIS AG.
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERN GES MBH.
XX
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX      WPI: 2000-097741/08.
XX      P-PDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
XX      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
XX      AAV58585, AAV58586, AAV58587, AAV58588, AAV58589, AAV58590, AAV58591,
XX      AAV58592, AAV58593, AAV58594.
XX
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy
XX
XX      Claim 14, Page 87-104; 174pp; English.
XX
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-58578) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiply drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
XX
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX      Query Match 100.0%; Score 4233; DB 21; Length 68750;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      1 ATGACGATCAATCAGCTTCTGAACAGAGCTCGACACAGGATATCAAGTGGCGCGAT 60
XX      11872 ATGACGATCAATCAGCTTCTGAACAGAGCTCGACACAGGATATCAAGTGGCGCGAT 11931
XX
XX      61 GGGGAGCGCCTTCAGATACAGGCGCCCAAGAAACGCGCTGAACCCGACCTGCTCTCGA 120
XX      11932 GGGGAGCGCCTTCAGATACAGGCGCCCAAGAAACGCGCTGAACCCGACCTGCTCTCGA 11991
XX
XX      121 ATCTCCGACCAAAAGACAGATCTGACGATGCTCTGCTGAGATCTCCCGCAATCC 180
XX      11992 ATCTCCGACCAAAAGACAGATCTGACGATGCTCTGCTGAGATCTCCCGCAATCC 12051
XX
XX      181 ATCTCCGACCAAAAGACAGATCTGACGATGCTCTGCTGAGATCTCCCGCAATCC 240
XX      12052 ATCTCCGACCAAAAGACAGATCTGACGATGCTCTGCTGAGATCTCCCGCAATCC 12111
XX
XX      241 TACTGCTGGCGCGGACAGAGCGCTTACGATCGTCCACAGGAGATCCAGCGCTATCGGAA 300
XX      12112 TACTGCTGGCGCGGACAGAGCGCTTACGATCGTCCACAGGAGATCCAGCGCTATCGGAA 12171
XX
XX      301 TAGACTGTAGAGATCTGAGCTGCGGAGCTGAGACCGCGCTTTCGGAAGATCGTCCGG 360
XX      12172 TAGACTGTAGAGATCTGAGCTGCGGAGCTGAGACCGCGCTTTCGGAAGATCGTCCGG 12231
XX
XX      361 CGGACGACATGCTTGGGCGCCACACAGCGTGCCTCCGACATGATCAGGTATGAGCTAAA 420

```

Db 12232 CGGACGACATGCTTGGGCCCCACAGCTGCCGACATGATGCAAGTATCGAGCTTAA 12291  
Qy 421 GTCCAGCCGCGACATCGAGATCATGATCGGCGGGCTCGACCCGGAGCAACGGGAAACG 480  
Db 12292 GTCCAGCCGCGACATCGAGATCATGATCGGCGGGCTCGACCCGGAGCAACGGGAAACG 12351  
Qy 481 AGGCTCGTGTGCTTGCAGATGCGATGTCGACCCGATCTATGACACCGAGCGCCCTCG 540  
Db 12352 AGGCTCGTGTGCTTGCAGATGCGATGTCGACCCGATCTATGACACCGAGCGCCCTCG 12411  
Qy 541 CTCTATCACTGCTGTGCGGCTTGTGCTGACGAGCGGCAAAACCCGCTGTGTGCTCACTATC 600  
Db 12412 CTCTATCACTGCTGTGCGGCTTGTGCTGACGAGCGGCAAAACCCGCTGTGTGCTCACTATC 12471  
Qy 601 GATCTCATTAAGTTGACCTTAGGCGACCTGTCATCATCTTCAAGAGCTGGCTCAGCTTC 660  
Db 12472 GATCTCATTAAGTTGACCTTAGGCGACCTGTCATCATCTTCAAGAGCTGGCTCAGCTTC 12531  
Qy 661 TACGAAAGTCCGAGACCTCTCTCCCTGTCTGAGACTCTGTAACGGCATTTATGATCTC 720  
Db 12532 TACGAAAGTCCGAGACCTCTCTCCCTGTCTGAGACTCTGTAACGGCATTTATGATCTC 12591  
Qy 721 GCGCTGAGTCTCGCAAGAGTCTGAGCGGCATCAACGATCGATGATTTACTGGAACGG 780  
Db 12592 GCGCTGAGTCTCGCAAGAGTCTGAGCGGCATCAACGATCGATGATTTACTGGAACGG 12651  
Qy 781 CGCATCGCGAGCTCCGACCTCCGCGGAGCGCTCCGATGAAAGCGGATCCATCTACCCCTG 840  
Db 12652 CGCATCGCGAGCTCCGACCTCCGCGGAGCGCTCCGATGAAAGCGGATCCATCTACCCCTG 12711  
Qy 841 AAGAGATCCGCTTCCGCAACAGAGATGCTGCGCTCGGACTCTCGGGGCTGATTC 900  
Db 12712 AAGAGATCCGCTTCCGCAACAGAGATGCTGCGCTCGGACTCTCGGGGCTGATTC 12771  
Qy 901 AAGCGGCTGTGCGGAGAGCGGCGGCTGAACCCCGAGCGGCGTATCTGCGTTCATTTTCC 960  
Db 12772 AAGCGGCTGTGCGGAGAGCGGCGGCTGAACCCCGAGCGGCGTATCTGCGTTCATTTTCC 12831  
Qy 961 GAGGATGCGGCGCTGAGAGCGGAGCGGCGGTTTAAAGCTCAATAAAGCTCTTCAC 1020  
Db 12832 GAGGATGCGGCGCTGAGAGCGGAGCGGCGGTTTAAAGCTCAATAAAGCTCTTCAC 12891  
Qy 1021 CGGCTCCCCGTCATCCGCGCGTGAACGATATCAACGGGGACTTCACTGATGCTCTC 1080  
Db 12892 CGGCTCCCCGTCATCCGCGCGTGAACGATATCAACGGGGACTTCACTGATGCTCTC 12951  
Qy 1081 CTGACATCGACACCACTCGCGCAAGAGCTTTCGAAACAGCGGCTTAAGCGTATTCAGAG 1140  
Db 12952 CTGACATCGACACCACTCGCGCAAGAGCTTTCGAAACAGCGGCTTAAGCGTATTCAGAG 13011  
Qy 1141 CAGCTGTGGGAAGGAGATGATATCACTGCAACGATATGAGTTCACGAGAGGCC 1200  
Db 13012 CAGCTGTGGGAAGGAGATGATATCACTGCAACGATATGAGTTCACGAGAGGCC 13071  
Qy 1201 GCCCGGCTCTGGGAGATCCAAGAGGCGCATTTGTTCCCGTGTGCTCTACGAGCGGCTT 1260  
Db 13072 GCCCGGCTCTGGGAGATCCAAGAGGCGCATTTGTTCCCGTGTGCTCTACGAGCGGCTT 13131  
Qy 1261 AACGCAAGTCTGTGTGTCACTCTGTGCAAGAGCTCGAACTCCGCTGTACACAGC 1320  
Db 13132 AACGCAAGTCTGTGTGTCACTCTGTGCAAGAGCTCGAACTCCGCTGTACACAGC 13191  
Qy 1321 ACGGAGATCTCTCAGCTGTGTGTATCATCACTCTACAGACAGATGGGAGCTCTGTC 1380  
Db 13192 ACGGAGATCTCTCAGCTGTGTGTATCATCACTCTACAGACAGATGGGAGCTCTGTC 13251  
Qy 1381 CTCGGTGGGACATGTCGACGAGTGTTCGCGCGGACCTTCTGAGAGACATGCTCGAA 1440  
Db 13252 CTCGGTGGGACATGTCGACGAGTGTTCGCGCGGACCTTCTGAGAGACATGCTCGAA 13311  
Qy 1441 GCGTACGTGTTTTTCTCGGCGGCTCACTGAGAAACATGGGAGTAAACAGGTGCGCTGT 1500

Db 13312 GCGTACGTGTTTTTCTCGGCGGCTCACTGAGAACCATGGGATGAACAGTGGCTGT 13371  
Qy 1501 TCGCTTCGCTGCGCCAGCTTGAAGCGGCGGAGCGCAACCGACCGCTGTCG 1560  
Db 13372 TCGCTTCGCTGCGCCAGCTTGAAGCGGCGGAGCGGCAACCGACCGCTGTCG 13431  
Qy 1561 AGGAGCATACGCTGCAACGCGCTGTTCGCGCGGCTGAGCAGCTGCCATGACCTC 1620  
Db 13432 AGGAGCATACGCTGCAACGCGCTGTTCGCGCGGCTGAGCAGCTGCCATGACCTC 13491  
Qy 1621 GCCGTGTGTGCGGCGGCAAGACGCTCAACGAAAGCTTTGGCGGCTTCGCGGGA 1680  
Db 13492 GCCGTGTGTGCGGCGGCAACGCTCAACGAAAGCTTTGGCGGCTTCGCGGGA 13551  
Qy 1681 CTTGGCGCGGCTGCGCGAGCAAGGAGCGGCGGCAACATTTGTCGCGGTGTATG 1740  
Db 13552 CTTGGCGCGGCTGCGCGAGCAAGGAGCGGCGGCAACATTTGTCGCGGTGTATG 13611  
Qy 1741 GAGAAAGCTGGAGCAGGTTGTGCGGCTTCTCGCGTCTCGAGTCAGGCGGCTTAC 1800  
Db 13612 GAGAAAGCTGGAGCAGGTTGTGCGGCTTCTCGCGTCTCGAGTCAGGCGGCTTAC 13671  
Qy 1801 GTGCCGATCATGCTCCGACCTTACCGGCGGAGGCTATCCACTCTCTGATCATGTGAG 1860  
Db 13672 GTGCCGATCATGCTCCGACCTTACCGGCGGAGGCTATCCACTCTCTGATCATGTGAG 13731  
Qy 1861 GTAAAGCTCGTGTGACGACGACCTAGGCTGTGATGCAAACTGTATGGCGCGCGGATC 1920  
Db 13732 GTAAAGCTCGTGTGACGACGACCTAGGCTGTGATGCAAACTGTATGGCGCGGATC 13791  
Qy 1921 CAGCGGCTGCTCTGAGCGAGGCGGCGGCTCGAAGCGGAGCGGACCTTCGATATG 1980  
Db 13792 CAGCGGCTGCTCTGAGCGAGGCGGCGGCTCGAAGCGGAGCGGACCTTCGATATG 13851  
Qy 1981 CCAATTCAGACACTTGGATCTGCGGTATGTATCTTACACTGCGGATCCACAGGCTTG 2040  
Db 13852 CCAATTCAGACACTTGGATCTGCGGTATGTATCTTACACTGCGGATCCACAGGCTTG 13911  
Qy 2041 CCAAGGGGAGTATGATGATCATCGGGGAGCGGCTCAACACATCTGGAATCAACAGAG 2100  
Db 13912 CCAAGGGGAGTATGATGATCATCGGGGAGCGGCTCAACACATCTGGAATCAACAGAG 13971  
Qy 2101 CGCTTCGAATAGGAGCCCGGAGACAGGAGTGTGCGCTCTCTCGCTGAGCTTCGATTC 2160  
Db 13972 CGCTTCGAATAGGAGCCCGGAGACAGGAGTGTGCGCTCTCTCGCTGAGCTTCGATTC 14031  
Qy 2161 TCGGCTATGATGTGTTGCGGATCTTGTGCGGCGGCGGTATGATCTGTGTCGAGCGG 2220  
Db 14032 TCGGCTATGATGTGTTGCGGATCTTGTGCGGCGGCGGTATGATCTGTGTCGAGCGG 14091  
Qy 2221 TCCAAGCTGCGGATCCGCGGCACTTGGGAGAGTTGATCGAACGAGAGAGTGAACGGTG 2280  
Db 14092 TCCAAGCTGCGGATCCGCGGCACTTGGGAGAGTTGATCGAACGAGAGAGTGAACGGTG 14151  
Qy 2281 TGAACCTCGTGCAGGCGCTGATGCGATGCTCTGTCAGCATTTTGAAGGTCGCGCCGAT 2340  
Db 14152 TGAACCTCGTGCAGGCGCTGATGCGATGCTCTGTCAGCATTTTGAAGGTCGCGCCGAT 14211  
Qy 2341 TCGCTCGTATGCTCTGCGGCTTTCGCTGTGAGCGGCGCATGATCCCGTGGGCTTG 2400  
Db 14212 TCGCTCGTATGCTCTGCGGCTTTCGCTGTGAGCGGCGCATGATCCCGTGGGCTTG 14271  
Qy 2401 CTTGGCGAGCTCGAGGCACTATAGGCGCGGCGGTGTCGATGATCAACCTGGGGGCGGCAAC 2460  
Db 14272 CTTGGCGAGCTCGAGGCACTATAGGCGCGGCGGTGTCGATGATCAACCTGGGGGCGGCAAC 14331  
Qy 2461 GAAAGCTGATCTGCTCATCTGAGTACCCCGTGAAGAACGTCATCTGTGCGAGC 2520  
Db 14332 GAAAGCTGATCTGCTCATCTGAGTACCCCGTGAAGAACGTCATCTGTGCGAGC 14391  
Qy 2521 ATCCCTTACGCGCGCTGCGCTGCGCAACGACGTTTCCAGCTGTCTGATGAGCGCTCGAA 2580  
Db 14392 ATCCCTTACGCGCGCTGCGCTGCGCAACGACGTTTCCAGCTGTCTGATGAGCGCTCGAA 14451

QY 2581 CCGCGCCGCTCTGGTTCCGGGCACTCTAATTGGGGGTGGGCTTGGCACTGGGC 2640  
Db 14452 CCGCGCCGCTCTGGTTCCGGGCACTCTAATTGGGGGTGGGCTTGGCACTGGGC 14511  
QY 2641 TACTGCGGATGAAGAAGAGAGCGCAAGAGCTTCTGTGACCCCGAGACCGGGAG 2700  
Db 14512 TACTGCGGATGAAGAAGAGAGCGCAAGAGCTTCTGTGACCCCGAGACCGGGAG 14571  
QY 2701 CCGCTCTACAAGACCGGCGATCTGGGCGCTTACCTTCCGATGAAACATCGATTCTATG 2760  
Db 14572 CCGCTCTACAAGACCGGCGATCTGGGCGCTTACCTTCCGATGAAACATCGATTCTATG 14631  
QY 2761 GGGCGGAGAGCAACCAATCAAGCTTCCGATTAACCGCTTGAAGTCCGGGAAATCGAG 2820  
Db 14632 GGGCGGAGAGCAACCAATCAAGCTTCCGATTAACCGCTTGAAGTCCGGGAAATCGAG 14691  
QY 2821 GAAACGCTCAAGTCGATCCGAACGTACGCGAGCGGATGTTGTGCCGTGGGAAACGAC 2880  
Db 14692 GAAACGCTCAAGTCGATCCGAACGTACGCGAGCGGATGTTGTGCCGTGGGAAACGAC 14751  
QY 2881 GCGGCGAACAGCTCTTCTAAGCTTATGTGTCCGAGAGGCAACGAGACGCGTGC 2940  
Db 14752 GCGGCGAACAGCTCTTCTAAGCTTATGTGTCCGAGAGGCAACGAGACGCGTGC 14811  
QY 2941 GAGCGAGAGCGGAGCTTCAAGCCGAGCGGATCGAGCGAGAGCAACGCGCGAGAGG 3000  
Db 14812 GAGCGAGAGCGGAGCTTCAAGCCGAGCGGATCGAGCGAGAGCAACGCGCGAGAGG 14871  
QY 3001 GACGCGTTAGCGACCGGAGAGAGGTGCACTTCAAGCTTCAAGCTTCAAGCTTCAAGG 3060  
Db 14872 GACGCGTTAGCGACCGGAGAGAGGTGCACTTCAAGCTTCAAGCTTCAAGCTTCAAGG 14931  
QY 3061 GACCTTGACGAAAGCCGCTGTGATCTGACCGGAGAGAGTCCGCGAGAGCGCGGCTG 3120  
Db 14932 GACCTTGACGAAAGCCGCTGTGATCTGACCGGAGAGAGTCCGCGAGAGCGCGGCTG 14991  
QY 3121 GACGCTAGCGCGGTGCGGAGCGGATCGCAACGCTTCAAGCGCGCGAGTCCGCTTGT 3180  
Db 14992 GACGCTAGCGCGGTGCGGAGCGGATCGCAACGCTTCAAGCGCGCGAGTCCGCTTGT 15051  
QY 3181 GAGTTTGTGATTCCTGAGCTGCTTGAAGAGCGGATCGCAACGCGCGAGTCCGCTTGT 3240  
Db 15052 GAGTTTGTGATTCCTGAGCTGCTTGAAGAGCGGATCGCAACGCGCGAGTCCGCTTGT 15111  
QY 3241 AAATTCGTTATTCATTCGCGGAGAGCAAGTCCGCTGCAACCTTACGCTATGCTCA 3300  
Db 15112 AAATTCGTTATTCATTCGCGGAGAGCAAGTCCGCTGCAACCTTACGCTATGCTCA 15171  
QY 3301 TCCGCGCGATGAGGCGGTGAGAGCGGCTTATTAACAACCGCTTCAAGCAACGCT 3360  
Db 15172 TCCGCGCGATGAGGCGGTGAGAGCGGCTTATTAACAACCGCTTCAAGCAACGCT 15231  
QY 3361 TTGCTGAAGCTTCTCGATCAAGGATGAGCGGAGCGCACCTTTCGCAAAAATTCTGAC 3420  
Db 15232 TTGCTGAAGCTTCTCGATCAAGGATGAGCGGAGCGCACCTTTCGCAAAAATTCTGAC 15291  
QY 3421 GTGTTGATGAAGCGGCTTCAACCTTCTGTTCTGTGGGAGATGCAACGCTATGCTG 3480  
Db 15292 GTGTTGATGAAGCGGCTTCAACCTTCTGTTCTGTGGGAGATGCAACGCTATGCTG 15351  
QY 3481 CTGTATGATGATGCGTGTGAGAAATTTGCTGCTGAGAGCGGATATATGCGGAGCTG 3540  
Db 15352 CTGTATGATGATGCGTGTGAGAAATTTGCTGCTGAGAGCGGATATATGCGGAGCTG 15411  
QY 3541 CTGTATGATGATGCGCTTCTGCAACATCGGCTGTCTCGGTGGGCAATTCATTTT 3600  
Db 15412 CTGTATGATGATGCGCTTCTGCAACATCGGCTGTCTCGGTGGGCAATTCATTTT 15471  
QY 3601 GAAACAGTTTCGCGGCTTCTGCACTGCGACATTCGAGCTTTTACGTGACGCGATGCTG 3660  
Db 15472 GAAACAGTTTCGCGGCTTCTGCACTGCGACATTCGAGCTTTTACGTGACGCGATGCTG 15531

QY 3661 GGGCGGCGGATGAGACCGCGGAGTTCCAGGCTGTACGCTCGATCAGGATTCCTCACCG 3720  
Db 15532 GGGCGGCGGATGAGACCGCGGAGTTCCAGGCTGTACGCTCGATCAGGATTCCTCACCG 15591  
QY 3721 AGGCGCGGATGAGACCGCGGAGTTCCAGGCTGTACGCTCGATCAGGATTCCTCACCG 3780  
Db 15592 AGGCGCGGATGAGACCGCGGAGTTCCAGGCTGTACGCTCGATCAGGATTCCTCACCG 15651  
QY 3781 CCGGCTTCTTGAAGACCAAACTTACCGAGTATGATGCTTCAAGCTTCTGTGAGCTG 3840  
Db 15652 CCGGCTTCTTGAAGACCAAACTTACCGAGTATGATGCTTCAAGCTTCTGTGAGCTG 15711  
QY 3841 GATGCTTCCGCTACGCTCAACGAGCAAGGTCATGCTGAAGCCCTGCGAGCGGAG 3900  
Db 15712 GATGCTTCCGCTACGCTCAACGAGCAAGGTCATGCTGAAGCCCTGCGAGCGGAG 15771  
QY 3901 GATACCTGCTGCGCGGATTCGAGGAGCAACGCGCGCAACGCGGAGCTTGGAGAGATC 3960  
Db 15772 GATACCTGCTGCGCGGATTCGAGGAGCAACGCGCGCAACGCGGAGCTTGGAGAGATC 15831  
QY 3961 CTGCTGCGGCTGTATCGGAGAGTCTCGGCTGTGAGGTGTGAGCTTCAAGCTTTC 4020  
Db 15832 CTGCTGCGGCTGTATCGGAGAGTCTCGGCTGTGAGGTGTGAGCTTCAAGCTTTC 15891  
QY 4021 GTGATCTTGTGAGATTCGATTCATTCATTCGCTTCAAGAGAGCTTGGAGAGAG 4080  
Db 15892 GTGATCTTGTGAGATTCGATTCATTCATTCGCTTCAAGAGAGCTTGGAGAGAG 15951  
QY 4081 CTGATTAAGAGATTCGATTCATTCATTCGCTTCAAGAGAGCTTGGAGAGAG 4140  
Db 15952 CTGATTAAGAGATTCGATTCATTCATTCGCTTCAAGAGAGCTTGGAGAGAG 16011  
QY 4141 TCCGCTTTCGCGGAGATTCGAGAGATCTAAGTACGCGCGGAGATGAGAGAGAG 4200  
Db 16012 TCCGCTTTCGCGGAGATTCGAGAGATCTAAGTACGCGCGGAGATGAGAGAGAG 16071  
QY 4201 GAGTTTCGCGGAGAGAGAGAGAGAGTGAAGCTTAA 4233  
Db 16072 GAGTTTCGCGGAGAGAGAGAGAGAGTGAAGCTTAA 16104

RESULT 2  
ID AAA29349 strand: DNA; 71989 BP.  
XX AAA29349;  
AC XX  
XX  
DT 12-SBP-2000 (first entry)  
XX  
XX  
DE Sorangium cellulosum epothione synthase operon genomic DNA.  
XX  
XX  
KW Epothione; polyketide synthase; epok; epob; epoc; epod; epoe; epof;  
KW epol; epok; P450 epoxidase; ORF A; ORF B; promoter; enhancer; anti-fungal;  
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.  
OS Sorangium cellulosum.  
XX  
XX  
FH Key  
FH CDS 3..992  
FT /tag= a  
FT /label= ORF A  
FT /product= trianposase  
FT /note= "not part of the PKS"  
FT 989..1501  
FT CDS  
FT /tag= b  
FT /label= ORF B  
FT /product= trianposase  
FT /note= "not part of the PKS"  
FT 1998..6263  
FT CDS  
FT /tag= c  
FT /label= epok\_gene  
FT /note= "encodes the loading domain"  
FT misc\_RNA 2031..3548



```
FT FT /*tag= d
FT FT /note= "encodes ketide synthase (KS-Y) of the loading
FT FT domain"
FT FT 3621..4661
FT FT /*tag= e
FT FT /note= "encodes acyl transferase (AT) of the loading
FT FT domain"
FT FT 4917..5810
FT FT /*tag= f
FT FT /note= "encodes enoyl reductase (ER) of the loading
FT FT domain, potentially involved in formation of the
FT FT thiazole moiety"
FT FT 5856..6155
FT FT /*tag= g
FT FT /note= "encodes acyl carrier protein (ACP) of the loading
FT FT domain"
FT FT 6260..10493
FT FT /*tag= h
FT FT /label= epob_gene
FT FT /note= "encodes module 1, the NRPS module"
FT FT 2031..3548
FT FT /*tag= i
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT 2031..3548
FT FT /*tag= j
FT FT /note= "encodes condensation domain C2 of the NRPS
FT FT module"
FT FT 6861..6887
FT FT /*tag= k
FT FT /note= "encodes heterocyclization signature sequence"
FT FT 6861..6887
FT FT /*tag= l
FT FT /note= "encodes condensation domain C4 of the NRPS
FT FT module"
FT FT 7358..7366
FT FT /*tag= m
FT FT /note= "encodes condensation domain C7 (partial) of the
FT FT NRPS module"
FT FT 7898..7921
FT FT /*tag= n
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT 7898..7921
FT FT /*tag= o
FT FT /note= "encodes adenylation domain A1 of the NRPS module"
FT FT 8261..8308
FT FT /*tag= p
FT FT /note= "encodes adenylation domain A3 of the NRPS module"
FT FT 8411..8422
FT FT /*tag= q
FT FT /note= "encodes adenylation domain A4 of the NRPS module"
FT FT 8861..8905
FT FT /*tag= r
FT FT /note= "encodes adenylation domain A6 of the NRPS module"
FT FT 8966..8983
FT FT /*tag= s
FT FT /note= "encodes adenylation domain A7 of the NRPS module"
FT FT 9090..9179
FT FT /*tag= t
FT FT /note= "encodes adenylation domain A8 of the NRPS module"
FT FT 9183..9992
FT FT /*tag= u
FT FT /note= "encodes oxidation region for forming thiazole"
FT FT 10121..10138
FT FT /*tag= v
FT FT /note= "encodes adenylation domain A10 of the NRPS
FT FT module"
FT FT 10261..10306
FT FT /*tag= w
FT FT /note= "encodes thiolation domain (PCP) of the NRPS
FT FT module"
FT FT 10639..16137
FT FT /*tag= x

FT FT /label= epoc_gene
FT FT /note= "encodes module 2"
FT FT 10654..12033
FT FT /*tag= y
FT FT /note= "encodes KS2, the KS domain of module 2"
FT FT 12250..13287
FT FT /*tag= z
FT FT /note= "encodes AT2, the AT domain of module 2"
FT FT 13327..13899
FT FT /*tag= aa
FT FT /note= "encodes dehydratase (DH) 2, the DH domain of
FT FT module 2"
FT FT 14962..15756
FT FT /*tag= ab
FT FT /note= "encodes ketoreductase (KR) 2, the KR domain of
FT FT module 2"
FT FT 15763..16008
FT FT /*tag= ac
FT FT /note= "encodes ACP2, the ACP domain of module 2"
FT FT 16134..37907
FT FT /*tag= ad
FT FT /label= epod_gene
FT FT /note= "encodes modules 3-6"
FT FT 16425..17606
FT FT /*tag= ae
FT FT /note= "encodes KS3"
FT FT 17817..18857
FT FT /*tag= af
FT FT /note= "encodes AT3"
FT FT 19581..20396
FT FT /*tag= ag
FT FT /note= "encodes KR3"
FT FT 20424..20642
FT FT /*tag= ah
FT FT /note= "encodes ACP3"
FT FT 20706..22082
FT FT /*tag= ai
FT FT /note= "encodes KS4"
FT FT 22296..23336
FT FT /*tag= aj
FT FT /note= "encodes AT4"
FT FT 24069..24647
FT FT /*tag= ak
FT FT /note= "encodes KR4"
FT FT 24867..25151
FT FT /*tag= al
FT FT /note= "encodes ACP4"
FT FT 25203..26576
FT FT /*tag= am
FT FT /note= "encodes KS5"
FT FT 26793..27883
FT FT /*tag= an
FT FT /note= "encodes AT5"
FT FT 27966..28574
FT FT /*tag= ao
FT FT /note= "encodes DHS"
FT FT 29433..30287
FT FT /*tag= ap
FT FT /note= "encodes ER5"
FT FT 30321..30869
FT FT /*tag= aq
FT FT /note= "encodes KR5"
FT FT 31077..31373
FT FT /*tag= ar
FT FT /note= "encodes ACP5"
FT FT 31440..32807
FT FT /*tag= as
FT FT /note= "encodes KS6"
FT FT 33018..34067
FT FT /*tag= at
FT FT /note= "encodes AT6"
FT FT 34107..34676
FT FT /*tag= au
```

FT /note="encodes DH6"  
FT misc\_RNA 35760..36641  
FT /tag= av  
FT /note="encodes ER6"  
FT misc\_RNA 36705..37256  
FT /tag= aw  
FT /note="encodes KR6"  
FT misc\_RNA 37470..37769  
FT /tag= ax  
FT /note="encodes ACP6"  
FT CDS 37912..49308  
FT /tag= ay  
FT /label= epor gene  
FT /note="encodes modules 7 and 8"  
FT misc\_RNA 38014..39375  
FT /tag= az  
FT /note="encodes KS7"  
FT misc\_RNA 39589..40626  
FT /tag= ba  
FT /note="encodes AT7"  
FT misc\_RNA 41341..41922  
FT /tag= db  
FT /note="encodes KR7"  
FT misc\_RNA 42181..42423

Query Match 97.2%; Score 4114.6; DB 21; Length 71989;  
Best Local Similarity 98.3%; Pired. No. 0;

Matches 4159; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGACATCAATAGCTTCTGAACGAGCTGACACAGGATATCAAGCTGGGGCGAT 60  
DB 6260 ATGACATCAATAGCTTCTGAACGAGCTGACACAGGATATCAAGCTGGGGCGAT 6319  
QY 61 GGGGAGCGCTCCAGATACAGGCCGCCAAGAACGCTTGAACCGAACCCTGCTCGTCA 120  
DB 6320 GGGGAGCGCTCCAGATACAGGCCGCCAAGAACGCTTGAACCGAACCCTGCTCGA 6379  
QY 121 ATCTCCGAGCAAAAGACAGATCTTGAAGATCTCTGAGAGATCTCCCGGAGATTC 180  
DB 6380 ATCTCCGAGCAAAAGACAGATCTTGAAGATCTCTGAGAGATCTCCCGGAGATTC 6439  
QY 181 ATCGTCCCGCCCGAGCGGAGCGGAGCGCTCGCTTCTCTCAGACATCCAAAGATCC 240  
DB 6440 ATCGTCCCGCCCGAGCGGAGCGGAGCGCTCGCTTCTCTCAGACATCCAAAGATTC 6499  
QY 241 TACTGCTGGGCGGAGCAGGAGCGTTTACGCTCCGAGCGGATCCAGCTTATCCGAA 300  
DB 6500 TACTGCTGGGCGGAGCAGGAGCGTTTACGCTCCGAGCGGATCCAGCTTATCCGAA 6559  
QY 301 TAGCACTGTAACGATCTTCCAGCTGCGAGGCTGAGCGCGCTTTGGGAAAGTCGTCG 360  
DB 6560 TAGCACTGTAACGATCTTCCAGCTGCGAGGCTGAGCGCGCTTTGGGAAAGTCGTCG 6619  
QY 361 CGGACGACATGCTTGGGGCCACACGCTGCCGACATGATGACGATGATGAGCTTAA 420  
DB 6620 CGGACGACATGCTTGGGGCCACACGCTGCCGACATGATGAGCTTAA 6679  
QY 421 GTGAGCGCGACATGAGATCATGATCTGCGGGGCTGACCGGAGCACACGGAAGCG 480  
DB 6680 GTGAGCGCGACATGAGATCATGATCTGCGGGGCTGACCGGAGCACACGGAAGCG 6739  
QY 481 AGGCTGCTGCTTGGGAGATGCGATGTCGACCGCATATGACACCGAGCCCTCCG 540  
DB 6740 AGGCTGCTGCTTGGGAGATGCGATGTCGACCGCATATGACACCGAGCCCTCCG 6799  
QY 541 CTCTATACGTCGTCGCGCTTGGGAGAGGAGCGGAAACCGCTCTGCTCAGTATC 600  
DB 6800 CTCTATACGTCGTCGCGCTTGGGAGAGGAGCGGAAACCGCTCTGCTCAGTATC 6859  
QY 601 GATCTCATTAACGTTGACCTGAGCAGCTGTCATCATCTTCAAGAGCTGGCTCAGCTTC 660  
DB 6860 GATCTCATTAACGTTGACCTGAGCAGCTGTCATCATCTTCAAGAGCTGGCTCAGCTTC 6919

QY 661 TAGCAAGATCCGAGACCTCTCTCCCTGTCGAGCTCTGTACCGGATTAATGATCTC 720  
DB 6920 TAGCAAGATCCGAGACCTCTCTCCCTGTCGAGCTCTGTACCGGATTAATGATCTC 6979  
QY 721 GCGCTGAGTCTCGAAGAGTCTGAGGCGATCAACGATGATGATTTACTGAAAGCG 780  
DB 6980 GCGCTGAGTCTCGAAGAGTCTGAGGCGATCAACGATGATGATTTACTGAAAGCG 7039  
QY 781 CGCATCGCGAGCTCCACCTTCCCGGAGCTTCCGATGAAAGCCGATCTCATCTACCTG 840  
DB 7040 CGCATCGCGAGCTCCACCTTCCCGGAGCTTCCGATGAAAGCCGATCTCATCTACCTG 7099  
QY 841 AAGGATCCGCTTCGAGCACAGGAGCAATGCTGCGTCCGATCCCTGAGGATGATTG 900  
DB 7100 AAGGATCCGCTTCGAGCACAGGAGCAATGCTGCGTCCGATCCCTGAGGATGATTG 7159  
QY 901 AAGCGCGCTGTCGAGGAGCGCGGCTGACCCCGAGCGGCGTATCTGCTGATTTTC 960  
DB 7160 AAGCGCGCTGTCGAGGAGCGCGGCTGACCCCGAGCGGCGTATCTGCTGATTTTC 7219  
QY 961 GAGGTATGAGGCGCTGAGGCGCGGAGCCCGGTTTACGCTTCAATACGCTTTCAAC 1020  
DB 7220 GAGGTATGAGGCGCTGAGGCGCGGAGCCCGGTTTACGCTTCAATACGCTTTCAAC 7279  
QY 1021 CGGCTCCCGTCAATCCGCGGAGTACAGATATCACCGGAGCTTACGTCGATGCTTC 1080  
DB 7280 CGGCTCCCGTCAATCCGCGGAGTACAGATATCACCGGAGCTTACGTCGATGCTTC 7339  
QY 1081 CTGGAATGACACCACTGCGACAAAGCTTGAACAGCGGCTTAAGGATTTCAAG 1140  
DB 7340 CTGGAATGACACCACTGCGACAAAGCTTGAACAGCGGCTTAAGGATTTCAAG 7399  
QY 1141 CAGCTTGGGAAAGCGATGATATCTGCAAGTAAAGCGTATGAGTCTCAGGAGAGG 1200  
DB 7400 CAGCTTGGGAAAGCGATGATATCTGCAAGTAAAGCGTATGAGTCTCAGGAGAGG 7459  
QY 1201 GCGCGGCTGCGGAGATCCAAAGAGCGATGTTCCCGTGTGCTCAGAGCGGCTT 1260  
DB 7460 GCGCGGCTGCGGAGATCCAAAGAGCGATGTTCCCGTGTGCTCAGAGCGGCTT 7519  
QY 1261 AACCAAGATGTTGTTGTTCACTCTGTTGACAGGCTTGAACCTCCGTTTACACAC 1320  
DB 7520 AACCAAGATGTTGTTGTTCACTCTGTTGACAGGCTTGAACCTCCGTTTACACAC 7579  
QY 1321 AGCAGACTCTCTGAGCTGCTGATGATATCAAGCTTCTAGACAGATGAGGACCTCTG 1380  
DB 7580 AGCAGACTCTCTGAGCTGCTGATGATATCAAGCTTCTAGACAGATGAGGACCTCTG 7639  
QY 1381 CTGCGGTGGAATGATGTCGACGAGTGTCCGCGGACCTTCTGGAAGACATGCTGAA 1440  
DB 7640 CTGCGGTGGAATGATGTCGACGAGTGTCCGCGGACCTTCTGGAAGACATGCTGAA 7699  
QY 1441 GCGTACGTGTTTTCTCCGGGCTCACTAGAGAACATGAGGTTGAACAGTCCGCTGT 1500  
DB 7700 GCGTACGTGTTTTCTCCGGGCTCACTAGAGAACATGAGGTTGAACAGTCCGCTGT 7759  
QY 1501 TCGCTTCCCGCTGCGACGTAAGCGGCGGAGCGGAAAGCGACCAACCGCTGCTG 1560  
DB 7760 TCGCTTCCCGCTGCGACGTAAGCGGCGGAGCGGAAAGCGACCAACCGCTGCTG 7819  
QY 1561 AGGAGCATACGCTGACGAGCTGTTGCGGCGGAGTGAACAGCTGCCATGAGCTC 1620  
DB 7820 AGGAGCATACGCTGACGAGCTGTTGCGGCGGAGTGAACAGCTGCCATGAGCTC 7879  
QY 1621 GCGTGTGTGCGCGCGCAAGACGCTCAGTACGAAGCTTTTCGCGCTTTCGCGCGA 1680  
DB 7880 GCGTGTGTGCGCGCGCAAGACGCTCAGTACGAAGCTTTTCGCGCTTTCGCGCGA 7939  
QY 1681 CTGGCGCGCGCTGCGACAGAGGAGGACCGCCGGAACAATGTTGCGCGGTGGTGA 1740  
DB 7940 CTGGCGCGCGCTGCGACAGAGGAGGACCGCCGGAACAATGTTGCGCGGTGGTGA 7999  
QY 1741 GAGAAAGCTGAGAGAGGTTGCGCGGTTCTCGCGGTCTGAGTCAAGCGCGGCTTAC 1800

Db 8000 GAGAAAGCTGGAGACAGGTTGTCGCGGTTCTCGCGTGTGAGTCAAGCCGCGCTAC 8059  
Qy 1801 GGGCGATCGATGCGGACCTACCGGCGAGCGGTATCCACTACTCTCGATATATGTAG 1860  
Db 8060 GGGCGATCGATGCGGACCTACCGGCGAGCGGTATCCACTACTCTCGATATATGTAG 8119  
Qy 1861 GTAAAGCTGTGTCTGACGAGCCATGAGTGAAGCAACTGTCAATGAGCCGCGGAGATC 1920  
Db 8120 GTAAAGCTGTGTCTGACGAGCCATGAGTGAAGCAACTGTCAATGAGCCGCGGAGATC 8179  
Qy 1921 CAGCGGCTGTGTGTGAGCGAGCCGCGGTGAAAGCGAGCGAGCCAGCTTCGATGATG 1980  
Db 8180 CAGCGGCTGTGTGTGAGCGAGCCGCGGTGAAAGCGAGCGAGCCAGCTTCGATGATG 8239  
Qy 1981 CCCATTCAAGACCTTCGAGTCTCGGTATGTCACTCACTCGGAGTCCACAGGTTG 2040  
Db 8240 CCCATTCAAGACCTTCGAGTCTCGGTATGTCACTCACTCGGAGTCCACAGGTTG 8299  
Qy 2041 CCCAAGGGGATGATGATCGATCGAGGTCGCTCAACACCATCTCGAGATCAACGAG 2100  
Db 8300 CCCAAGGGGATGATGATCGATCGAGGTCGCTCAACACCATCTCGAGATCAACGAG 8359  
Qy 2101 CGCTTGAATAGGAGCCCGGAGACAGAGTGTGCGGCTCTCTCGGTAGCTTCATCTC 2160  
Db 8360 CGCTTGAATAGGAGCCCGGAGACAGAGTGTGCGGCTCTCTCGGTAGCTTCATCTC 8419  
Qy 2161 TCGGCTAATGATGTGTGCGGATCTCTGCGGCGGCGGTACATGATGTGCGGACCG 2220  
Db 8420 TCGGCTAATGATGTGTGCGGATCTCTGCGGCGGCGGTACATGATGTGCGGACCG 8479  
Qy 2221 TCCAAGCTGCGGATCCGCGGATTTGGGAGAGTTGATGAAAGAGAGAGTGAACGAG 2280  
Db 8480 TCCAAGCTGCGGATCCGCGGATTTGGGAGAGTTGATGAAAGAGAGAGTGAACGAG 8539  
Qy 2281 TGAAGCTGCGTCCGCGCTGATCGGATGCTGTGAGACATTTTGAAGGTCGCCCAT 2340  
Db 8540 TGAAGCTGCGTCCGCGCTGATCGGATGCTGTGAGACATTTTGAAGGTCGCCCAT 8599  
Qy 2341 TCGGCTGAGGATCTGTGCGGCTTTTCTGCTGAGACGCGGATGATCCCGGTGGGCTG 2400  
Db 8600 TCGGCTGAGGATCTGTGCGGCTTTTCTGCTGAGAGCGGACTGATCCCGGTGGGCTG 8659  
Qy 2401 CTTGCGAGCTCAGGCGCATCAGGCGCGGCGGTGTGAGTCAAGCTGCGGCGGCGCAC 2460  
Db 8660 CTTGCGAGCTCAGGCGCATCAGGCGCGGCGGTGTGAGTCAAGCTGCGGCGGCGCAC 8719  
Qy 2461 GAAAGCTGATGTGTCATCGGATACCCGTTGAGGAACGTGACCTATCTGTGGCGAGC 2520  
Db 8720 GAAAGCTGATGTGTCATCGGATACCCGTTGAGGAACGTGACCTATCTGTGGCGAGC 8779  
Qy 2521 ATCCCTACGCGCGCTCCGCTGCGCAACAGAGTTTCCAGTGTGATGAGCGCTCGAA 2580  
Db 8780 ATCCCTACGCGCGCTCCGCTGCGCAACAGAGTTTCCAGTGTGATGAGCGCTCGAA 8839  
Qy 2581 CCGCGCGGATGTGAGTTCCGCGGCAACTCTACATTTGGCGGATCGGAGTGGCACTGGAC 2640  
Db 8840 CCGCGCGGATGTGAGTTCCGCGGCAACTCTACATTTGGCGGATCGGAGTGGCACTGGAC 8899  
Qy 2641 TACTGCGCGATGAGAGAGAGACGCGCAAGACTTCTCTGTGACACCCCGAGACCGGAG 2700  
Db 8900 TACTGCGCGATGAGAGAGAGACGCGCAAGACTTCTCTGTGACACCCCGAGACCGGAG 8959  
Qy 2701 CGGCTCTACAGACCGCGGATTTGGGCGGCTACCTGCGGATGAGAAACTGAGTTCAATG 2760  
Db 8960 CGGCTCTACAGACCGCGGATTTGGGCGGCTACCTGCGGATGAGAAACTGAGTTCAATG 9019  
Qy 2761 GGGCGTGAAGCAACCAATCAAGCTTCCGAGATACCGGATGAGCTCGGAGAAATCGAG 2820  
Db 9020 GGGCGGAGAGCAACCAATCAAGCTTCCGAGATACCGGATGAGCTCGGAGAAATCGAG 9079  
Qy 2821 GAAAGCTCAAGTGCATCGAAGTACGCGACGCGGTGATTTGTGCGGTGGAGACAGC 2880

Db 9080 GAAAGCTCAAGTGCATCCGAAGTACGAGACGCGGTGATGTGTCCTCGGAGACAGC 9139  
Qy 2881 GGGCGAACAAGCTCTCTTACGCTATGTGTCGCGAGGAGCAACGAGACGCGTGC 2940  
Db 9140 GGGCGAACAAGCTCTCTTACGCTATGTGTCGCGAGGAGCAACGAGACGCGTGC 9199  
Qy 2941 GAGCAGACGCGAGCTTCAAGACCGAGGATTCAGCGGAGAGCAACAGCGCGCAAGCG 3000  
Db 9200 GAGCAGACGCGAGCTTCAAGACCGAGGATTCAGCGGAGAGCAACAGCGCGCAAGCG 9259  
Qy 3001 GAGCGCTTGAAGCAGCGGAGAGGATTCAGAGTTCAGAGCTTCGACAGGATCCGAGG 3060  
Db 9260 GAGCGATTTGAGCAGCGGAGAGGATTCAGAGTTCAGAGCTTCGACAGGATCCGAGG 9319  
Qy 3061 GAGCTTGAAGGAGGAGCGGTGTGATGACCGGAGCGAGATTCGCGGAGGCGGAGCTG 3120  
Db 9320 GATCTGAGCGAAGACCGGTGTGATGACCGGAGCGAGATTCGCGGAGGCGGAGCTG 9379  
Qy 3121 GACGTCTACGCGCGTCCGATGAGCGTCCGAAGCTTCTGAGGCGCCGATTCGTTGTT 3180  
Db 9380 GACGTCTACGCGCGTCCGATGAGCGTCCGAAGCTTCTGAGGCGCCGATTCATTTGTT 9439  
Qy 3181 GAGTTTGTGATTCGAGCTCTGAGCTGTTGAGCAGCGTGAAGCCGAGCGGCGCACTTCC 3240  
Db 9440 GAAATTCGCGCATTCGAGCTGCTGAGCAGCGTGAAGCCGAGCGGCGGCGCTTCC 9499  
Qy 3241 AAATTCGCTTATCCATCGGCGGAGCAGCGTACCGGAGCAACCTACGCGATGCAAA 3300  
Db 9500 AAATTCGCTTATCCATCGGCGGAGCAGCGTACCGGAGCAACCTACGCGATGCAAA 9559  
Qy 3301 TCCGCGCGCATCGAGGCGTGAAGAGGAGCTTCTATTTACACCCGTTGAGACCGT 3360  
Db 9560 TCCGCGCGCATCGAGGCGTGAAGAGGAGCTTCTATTTACACCCGTTGAGACCGT 9619  
Qy 3361 TTGCTGAAGCTCTCGATCAAGGATGAGAGCGGAGCGGACGTTGGGAAAATTCCAG 3420  
Db 9620 TTGCTGAAGGCTCTCGATCAAGGATGAGAGCGGAGCGGACGTTGGGAAAATTCCAG 9679  
Qy 3421 GTGTTGATGAAGGCGGCTTCAACCTGTTTGGGAGAGATGAGCGCATGAGTGC 3480  
Db 9680 GTGTTGATGAAGGCGGCTTCAACCTGTTTGGGAGAGATGAGCGCATGAGTGC 9739  
Qy 3481 CTGATGATGATGCTGCGAGAAATTTTCTGCTGAGGCGGAGATATATGCGGAGCTC 3540  
Db 9740 CTGATGATGATGCTGCGAGAAATTTTCTGCTGAGGCGGAGATATATGCGGAGCTC 9799  
Qy 3541 CTGATGAGCAGGCGCTTCTCTGCAACATCGGCGTCTGTCCGTTGGGCAATTCAATTT 3600  
Db 9800 CTGATGAGCAGGCGCTTCTCTGCAACATCGGCGTCTGTCCGTTGGGCAATTCAATTT 9859  
Qy 3601 GAAAGGTTGCGGCGGTTCTGACCTGCGAGATTCGAGCGTTTACGTGACGCGCATGCTG 3660  
Db 9860 GAAAGGTTGCGGCGGTTCTGACCTGCGAGATTCGAGCGTTTACGTGACGCGCATGCTG 9919  
Qy 3661 GCGCGGCGGATGAGACCGCGGCAAGTTCAGGTTCTGAGCTTCGAGCAAGTTCTCACCG 3720  
Db 9920 GCGCGGCGGATGAGACCGCGGCAAGTTCAGGTTCTGAGCTTCGAGCAAGTTCTCACCG 9979  
Qy 3721 AGGCGCGCAGACGCGGCGGCTTCCGCGCGGAGCAGCACTTCCGATATGCTT 3780  
Db 9980 AGGCGCGCAGACGCGGCGGCTTCCGCGCGGAGCAGCACTTCCGATATGCTT 10039  
Qy 3781 CGGACCTTCTGAGAGCAACCACTACCGAGTACATGAGGCTTACGATTTGTGAGGCTC 3840  
Db 10040 CGGACCTTCTGAGAGCAACCACTACCGAGTACATGAGGCTTACGATTTGTGAGGCTC 10099  
Qy 3841 GATGCTTCCGCTGACGTTCCAGAGGATTCAGAGGCTTCCGAGCGGAGG 3900  
Db 10100 GATGCTTCCGCTGACGTTCCAGAGGATTCAGAGGCTTCCGAGCGGAGG 10159  
Qy 3901 GATACCTGTGCGCGGCAATTCGAGGAGCAACGCGGCGCAGCGGAGCTTGGAGAGATC 3960  
Db 10160 GATACCTGTGCGCGGCAATTCGAGGAGCAACGCGGCGCAGCGGAGCTTGGAGAGATC 10219

Oy		3961	CTCGTCGGGATGTCTGCACGGAGAGTGTGGGCTGGAGGTGGTTCCGAGCAACTTC	4020
Dd		10220	CTCGTTGGCGATCTGTAACGGAGAGTCTCGGGCTGGAGGTGGTTGGCTCCAGCAGAACTTC	10278
Oy		4021	GTCGATCTTGTTGGTGACATCGATTACATCGTTTGCATGAGAAGCCTGTTGCAGAAAGAG	4080
Dd		10280	GTCGATCTTGTTGGTGCAATCATGATTACATCGTTTGCATGAGAAGTGTGTTGCAGAAAGAG	10333
Oy		4081	CTGGAATAGGAGATTCGCCATATACCGAGTTGTTCCAGTACCAGAACCTCGGCTGCTGGGG	4140
Dd		10340	CTGGAATAGGAGATTCGCCATATACCGAGTTGTTCCAGTACCAGAACCTCGGCTGCTGGGG	10399
Oy		4141	TCCGATTGGCCCGCAGACTCGAGAGATCTAGATATAGAGGGCCGAACATGACAGACGAGTG	4200
Dd		10400	TCCGATTGGCCCGCAGACTCGAAGATCTTAGAGCAGAGCCGACCAATCATGAGACCAAGTG	10455
Oy		4201	GAGGTTCCGCGCAAGGAGGAGAGAGTAGCTAA	4233
Dd		10460	GAGGTTCCGCGCAAGGAGGAGAGAGTAGCTAA	10492
<hr/>				
RESULT 3				
AAAS8471				
ID	AAAS8471	standard; DNA; 58857 BP.		
XX	AAAS8471;			
XX				
DT	31-OCT-2000	(first entry)		
XX				
DE	Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.			
XX				
KW	BLM gene cluster; bleomycin gene cluster; polyketide metabolite;			
KM	bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;			
XX	thiazoline; dithiazoline; microbial metabolite; sugar; ss.			
OS	Streptomyces verticillius.			
XX				
FH				
Key	Location/Qualifiers			
FT	223..564			
FT	/tag= a			
FT	/transl_except= (pos: 1..3, aa: Met)			
FT	/note= "ORF 30; encodes AAB07556"			
CDS	561..2309			
FT	/tag= b			
FT	/transl_except= (pos: 1..3, aa: Met)			
FT	/note= "ORF 29; encodes AAB07557"			
CDS	2767..3486			
FT	/tag= c			
FT	/note= "ORF 28; encodes AAB07558"			
FT	3527..5593			
FT	/tag= d			
FT	/transl_except= (pos: 1..3, aa: Met)			
FT	/note= "ORF 27; encodes AAB07559"			
CDS	5806..12294			
FT	/tag= e			
FT	/note= "ORF 26; encodes AAB07560"			
FT	12291..15491			
FT	/tag= f			
FT	/note= "ORF 25; encodes AAB07561"			
CDS	15488..21013			
FT	/tag= g			
FT	/note= "ORF 24; encodes AAB07562"			
FT	21010..24666			
FT	/tag= h			
FT	/transl_except= (pos: 1..3, aa: Met)			
FT	/note= "ORF 23; encodes AAB07563"			
CDS	24663..32690			
FT	/tag= i			
FT	/note= "ORF 22; encodes AAB07564"			
FT	32893..34830			
FT	/tag= j			
FT	/note= "ORF 21; encodes AAB07565"			
FT				

FT	CDS	34827..35804
FT		/+tag= x
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 20; encodes AAB07566"
FT	CDS	35818..37302
FT		/+tag= 1
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 19; encodes AAB07567"
FT	CDS	37299..39215
FT		/+tag= m
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 18; encodes AAB07568"
FT		39301..47181
FT		/+tag= n
FT	CDS	47178..49985
FT		/note= "ORF 17; encodes AAB07569"
FT		/+tag= o
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 16; encodes AAB07570"
FT	CDS	49982..51001
FT		/+tag= p
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 15; encodes AAB07571"
FT	CDS	50998..52386
FT		/+tag= q
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 14; encodes AAB07572"
FT	CDS	52383..52946
FT		/+tag= r
FT		/note= "ORF 13; encodes AAB07573"
FT	CDS	53018..54190
FT		/+tag= s
FT		/note= "ORF 12; encodes AAB07574"
FT	CDS	54187..55824
FT		/+tag= t
FT		/note= "ORF 11; encodes AAB07575"
FT	CDS	55821..56093
FT		/+tag= u
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 10; encodes AAB07576"
FT	CDS	56090..57586
FT		/+tag= v
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 9; encodes AAB07577"
FT	CDS	57583..58857
FT		/+tag= w
FT		/note= "ORF 8; encodes AAB07578"
XX		
XX	PN	W0200040704-A1.
XX	PD	13-JUL-2000.
XX		
PF		06-JAN-2000; 2000MO-USO0445.
XX		
PR		06-JAN-1999; 99US-0115435.
PR		05-FEB-1999; 99US-0118848.
PR		05-JAN-2000; 2000US-0447962.
XX		
PA		(REGC ) UNIV CALIFORNIA.
XX		
PI		Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
XX		
DR		WPI; 2000-465974/40.
DR		P-PSDB: AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR		AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR		AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
XX		AAB07576, AAB07577, AAB07578.
XX		
PT		New bleomycin gene cluster components useful for peptide and/or
PT		polyketide metabolites, especially bleomycin, production and for
PT		chemically modifying biological molecules -
XS		Claim 8; Page 97-136; 16zpp; English.



```

Db 44297 TCTTCACTCGGCTCCACCGGGAACCAAGGCGTCATGATCGACACCTCGGGCGG 44356
Qy 2075 TCAACCACTCTGACATCAACGAGCGCTTCGAAATAGGCCCGGAGACAGGCTCTG 2134
Db 44357 CCACACCCCTGGAAATGCTCAACCGCGCTTCGGGACACCGGCGGAGCGGCTCTG 44416
Qy 2135 CGCTCCCTCGGCTGACCTGATCTCGTCTATATATGTTGTTGGGATCTCGGGCGG 2194
Db 44417 CGCTCTCTCCCGACCTTCGACCTTCGCTTACACCTGTTGGCGTGTGGCCGCG 44476
Qy 2195 GCGGTACATCGTGTGTCGCGGACGCTCAAGCTGCGCATCCGGGCGCATTTGGGAGAGT 2254
Db 44477 GCGGCAACCGTGTGCTCTCCCGCCACGACCGCCGGGCGACCCCGGACACTGGGCGGAGC 44536
Qy 2255 TGATCGAAGAGAAAGTGAAGCTGCTGTAATCTGCTCGGCGCTGATGCGGATCTCG 2314
Db 44537 TGATCGGCGCGGCGGCTCAACCTGTGGAATCTCGCTCCGCGCTGGGACACCTCTCA 44596
Qy 2315 TCGAGCATTTTGAAGGCTCGCCCGGATTCGCTCGCTAGTCTGCGGCGCTTTCGCTGA 2374
Db 44597 CCGAGTACCGCCGAGCGCTTCGCTCCGACGCTTCGCGACCTTCGCGGCGGCTCTCTCA 44656
Qy 2375 GCGGCGACTGATCCCGGCTGCGCTGCTGCGAGCTCCAGGCGAT---CAGGCGCGGCG 2431
Db 44657 GCGGCGACTGATCCCGCTCTCGACTGCGGACCGGATCCGCGCTCTGCTCGCCCGGCG 44716
Qy 2432 TGTGCGTATCAGCTCGGCGGCGGCGACCGAGCGTGCATCTGCTCATCGGCTACCCG 2491
Db 44717 CCACCGTATGAGCTTCGCGGCGCGGACCGAACCCTTCATCTGCTGCTGTGTAAGAGA 44776
Qy 2492 TGAAGAACGTCGACCTATGCTGGGCGAGCATCCCTTAAGCGCGCTCGGCTCGGACACGA 2551
Db 44777 TCGGGAAGGTGACGAGGCGTGAAGACATCCCTTAAGCGACCCCATAGGCCAACACG 44836
Qy 2552 CGTTCACGCTGCTGATGAGGCGCTGGAACCGCGCGCTGCTGCTTCGCGGCGCACTCT 2611
Db 44837 GGTGGAAGGTCTCTGACGAGACCTGCGCGCGCTCGGCGCGACCTGGTGCCGCGGAGCTGT 44896
Qy 2612 ACATTGGCGGGGCTGGGCTGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2671
Db 44897 ACATCGGCGGACCGGCGCTGCGCAAGGCTATGCGGCGGACCGCGAACAAGACCTCTCG 44956
Qy 2672 GCTTCTCTGTCACCCCGGAGACCGGGGAGCGCTTACAAAGACGGGATCTGGGCGCT 2731
Db 44957 GCTTCTCTCTGTCACCCCGGAGACCGGGGAGCGCTTACAAAGACGGGATCTGGGCGCT 45016
Qy 2732 ACCTGCGCGATGGAACATCGAGTTCAATGGGCGTGAAGACAAATCAAGCTTCGCG 2791
Db 45017 ACCTGCGCGACGCGACGCTGGAATCTCTGGGCGGCGAGACGACAGTGAATCGGGG 45076
Qy 2792 GATACCGCGTTGAGCTCGGGGAAATGAGGAAACGCTCAAGTGCATCCGAAGTACGCG 2851
Db 45077 GATTCGCGGTGCAATCGGGGAGAGTGAAGGCGGCGCTTCGCGACATCGCGAGTGCAG 45136
Qy 2852 AGCGGATGATGTCGCGGTCGGAAGCGAGCGGGAACAAGTCTTCTAGACTATGCG 2911
Db 45137 CCGGCGGCTGATGCTGCACTGTAACCGCGGGGCGAGCGCGCTGCTGCTGCGCG 45196
Qy 2912 TCCCGG 2917
Db 45197 TACCGG 45202

```

RESULT 4  
ABA89198  
ID ABA89198 standard; DNA; 4305 BP.  
XX  
AC ABA89198;  
XX  
XX 11-FEB-2002 (first entry)  
XX  
DE Escherichia coli polynucleotide SEQ ID NO 940.  
XX

```

KW Escherichia coli; B2/D+A-; anti-inflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance; ds.
OS Escherichia coli.
PN WO200166572-A2.
XX
XX 13-SEP-2001.
PD
XX 12-MAR-2001; 2001WO-EP03445.
PF
XX 10-MAR-2000; 2000FR-0003145.
PR 02-FEB-2001; 2001FR-0001449.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
PI WPI; 2001-550253/61.
DR
XX
XX A library of DNA fragments of Escherichia coli strains for the
PT phylogenetic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
XX Example 6; Fig 6; 646pp; English.
PS
XX
XX The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA8577-ABA88729 and ABA89533)
CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenetic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicaemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
XX Sequence 4305 BP; 880 A; 1148 C; 1309 G; 968 T; 0 other;
SQ
Query Match 12.8%; Score 543.4; DB 22; Length 4305;
Best Local Similarity 48.5%; Pred.No. 4.9e-106;
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;
Qy 181 ATCGTCCCGCGCCGACGAGCGGCGAGCGCTTCCTCTCAGACATCCAAAGATCC 240
Db 241 ATATGTCACGACCTGAGCATGTATCAACCGTTCCATTAAGCGAGTTCAGCGCGCC 300
Qy 241 TACTGCTGGGCGGACAGAGCGTTTACGCTCCAGCGGATCCAGCGCTATGCGGA 300
Db 241 ATATGTCACGACCTGAGCATGTATCAACCGTTCCATTAAGCGAGTTCAGCGCGCC 300
Qy 301 TACTGCTGGGCGGACAGAGCGTTTACGCTCCAGCGGATCCAGCGCTATGCGGA 300
Db 301 TACTGCTGGGCGGACAGAGCGTTTACGCTCCAGCGGATCCAGCGCTATGCGGA 300
Qy 301 TACGATGACGATCTGACAGTCCGAGGCTGAGACCGCGCTTCGGAATTCGTCGG 360
Db 301 TACTGCTGGGCGGACAGAGCGTTTACGCTCCAGCGGATCCAGCGCTATGCGGA 300
Qy 358 TTTGACGTGAAACACTTTAATGTTAAGGCTTTTACCATGCGGTAAATGCGGTAGCT 417
Db 358 TTTGACGTGAAACACTTTAATGTTAAGGCTTTTACCATGCGGTAAATGCGGTAGCT 417
Qy 361 CCGACGACATGCTTCGGGCGGACAGCGCTCCCGGACATGATGATGAGCTTA 420
Db 418 CCGCATGAAATCTACGTGCGGGTATCTCCCGAGCGTACTCAGACATTTCTGGCGCA 477
Qy 421 GTGACGCG---CGACATCGATCATGATTCGCGGCGGCTGACCGGAGCACCGGGA 477
Db 478 GTGCGCGGATCATGATGAGAGCGGATGAGTGTCTTCTCCCTTAACGACGAAAC 537
Qy 478 GCGAGGCTGTGCTGTGAGATGCAATGTCGACCGCATTAATGACACGAGGCGCT 537
Db 538 GATGCTTATGATGCAATCGCATCGGCTGTGATCATGATCATCCGCAATGCTGG 597
Qy 538 CCGCTTATCATGATGCTGCTGCTGCTGAGACGAGCGCAACCGCTGCTGCTCACT 597

```

Db 538 CCGCTGTTGATTTAGTTATTCGGCTTGCACGGGCGCAACATGCGCTGTCATTTCAAGT 657  
Qy 538 ATGCATCTATTAACTGAGACCTAGGACGCTGTCCATCATCTTCAAGGACCTGGCTACG 657  
Db 538 CTGCATCTGCTATTTGCTGCTATCTGATATGCGACCTACAGAGAGATTGATGATG 717  
Qy 658 TTCTACGAAGATCCGAGACCTCTCTCCCTGTCTGAGACTCTGTAACCGCATTTATGTA 717  
Db 718 CTGTAACCGTAGACCCCATGCTGATGCTGCTGCTACCTGTTCTCTTTCTGTGATCAAG 777  
Qy 718 CTGCGGCTGAGATCTGCGAAGATCTGAGGCGCATCAACATCGATGATGATTAAGGAG 777  
Db 778 CAGGGGCTGTTGATGAGACGAGCTGATGACGCGATCAGGCTTATTTGGCA 837  
Qy 778 CCGGCATCGCGGATCCACCTCCGCGGACGCTTCGATGAGGCGGATCATCTAC 837  
Db 838 CCGGGGCTGCGCACTGATATGAGCCACCAACGCTGCCGTACAGGCGATTTGGCGCA 897  
Qy 838 CTGAAGAGATCCGCTTCCGGGCAACGAGCAATGCTGCGCTGCACTCTGGGAGTGA 897  
Db 898 CTGCTCGGATCAGGTTGATGCTGCGCGCTCATCGGCTGTGACCCCAACCTGGGAGTG 957  
Qy 898 TTGAACGCGCTGTGCGGGAGCGCGGCTGATCCCGAGCGGCGTATCTGCTGCAATT 957  
Db 958 CTGACCGGCTGCGCCCAACGCAACGATACCAAGACGCGCATTTGTTGACAGTCTTT 1017  
Qy 958 TCCGAGGATGAGGGGCTGAGCGCGAGCGCGGTTTACGCTCAACATTAAGCTCTTC 1017  
Db 1018 AGCCAAAGTGTGACAGTGTGAGCCTTAGCCGACGTTTACGCTCATCTGACGTTGTC 1077  
Qy 1018 AACCGGCTCCCGCTCATCCGCGCTGACATATCACCGGAGCTTCACTGCTGATGTC 1077  
Db 1078 AACCGCGCGAGAGTTACCCCAACGAGAGGAGTATTTGATTTTACCGCTGTACG 1137  
Qy 1078 CTCTGAGATGACACCACTGCGCAACAGAGCTTCAACAGCGGCTTAAGCTTCA 1137  
Db 1138 TTGCTGAATGTTTGTACACAGCAGCAGCTTATGCCCCCAACAGCTCAGGATTCAG 1197  
Qy 1138 GAGCAGCTGTGGGAGCGATGATCATCTGACAGCTGAAGCGGAT-----GAGATC 1188  
Db 1138 GTGCACTGTGGGAGAGATCTGCAACATCTCTTCAAGTGGATCCGCGCCACGAGAGG 1257  
Qy 1189 CAGCGAGAGCGCGCGGCTCTGAGGATCCAAACGAGCGCATTTCCCGTGTGCTC 1248  
Db 1258 CTGATCCATAGCGGCTGTTTCATGCGCGATGCGGCTGATTAACATGATGTTGGAT 1317  
Qy 1249 ACGAGCGGCTTAACACGAGATCTGTGTGATCACTGTTCAAGGCTCGAACTCG 1308  
Db 1318 ATCGACGGGAGACACTGCGCAAGACCTCGGACACAAACCGTTTACTCTGTGTCG 1377  
Qy 1309 GTGTACACAGACGACGATCTCTGACGCTGCTGATCATCACTGACAGACAT 1368  
Db 1378 GACGCCAATATTATCCCAACACCGAGGTGTGCTGATCACCAAGTATCAAGTTGGT 1437  
Qy 1369 GGGGACCTGCTCTGCGGTGGACATCTGCAAGAGTGTCCCGCCGACCTTCTGAC 1428  
Db 1438 GGGGAGTTGCAATTTCACTGGGACCGGCTGAGCACTGTTGATACACAGCTGTGAT 1497  
Qy 1429 GACATGCTGAAAGCTACGTTTCTCCGCGGCTCACTGAGAAACATGAGGATGA 1488  
Db 1488 CAGATGTTGTGCTTATGCTGATGCTGAGGCGGCTGTGTCATGCG-----G 1548  
Qy 1489 CAGGTGCGCTGTCTCCGCTGCGCATGAGAGCGGGCGAGCGCAACCGCAC 1548  
Db 1549 CAAATGTTGGGGGATTAATGTTCTGTGCGCTGCGCACCGTTAAGTACCGGTACG 1608  
Qy 1549 AACGCGCTGCTGAGCGAGCATCGCTGACGCGCTGTTGCGCGCGGCTGAGCAGCTG 1608  
Db 1609 CAGGCTCTGACACTACGCGCTTGTGTGACCATGATTACTGCTGACGACACTGACG 1668  
Qy 1609 CCCATGACGCTGCGGTGTGTGCGGCGCAAGACGCTCAAGTACGAAGCTTTTCGCGC 1668

Db 1669 CCACAGAACTGCGCTGATCAGTCTATCCGTGAATTAACCTATCCCACTGTGACG 1728  
Qy 1669 GCTTGGCGGACATTTGCGCGCGGCTGCGGACGAGGAGGACGCGCCCAACATTTGTC 1728  
Db 1729 GCGGCGATCATGTGCGCGCGGCTGTTAGCGCTGCGGCTGACATGCGACCGCGTG 1788  
Qy 1729 GCGGTGTGATGAGAAAGCTGGAGACGAGTGTGCGGTTCTTCGCGTCTGATGA 1788  
Db 1789 GCGGTGTGATGAGAAAGCTGGAGACGAGATTTGCGCGCTTACAGGACATTTACAGT 1848  
Qy 1789 GCGGCGCTTACGTGCGCATGATGCGCACTTACCGGAGGCGTATCACTACTCTC 1848  
Db 1849 GATGAGTGTATCTGCAAGTGTGCTGATGCTGATGCGGATGCGGCTGCTGCTG 1908  
Qy 1849 GATCATGTGAGATTAAGCTGCTGATGCGGACGATGCTGATGCGGATGCTGATG 1908  
Db 1909 ACGGTGGCGAGGTGCGGCTGATCAAGTACGACCGGCTCTACGCAATTTGAGCGCTG 1968  
Qy 1909 CCGCGGGGATTCAGCGGCTGCTGAGCGAGCGGCGTGAAGCGGACGCGACG 1968  
Db 1969 CTGCGCG-----TGCTGATCATGACGACGGAATGTGGAACGCGCTGCTG 2016  
Qy 1969 CCTCGATGATGCCATTCAGACACTTGGATCTGCGGTATGTATCTACACTTCGGA 2028  
Db 2017 CCGTTGCTGAAGTGTGCGGAGTGTACGAGATCTGGCTATATCATTTTCACTTCGCG 2076  
Qy 2029 TCCACAGGTTGCCCAAGGGGATGATGATCATCGGAGGTCGCTCAACACCATCTG 2088  
Db 2077 TCCACCGGTATCCGGAAGAGATGATGATCAACCGCTGCGGCTATACAGCTGGA 2136  
Qy 2089 GACATCAACGAGCGCTTGAATTAAGGCGCGGAGACAGGCTGTGCGCTCTCTGCTG 2148  
Db 2137 GACATCAACGAAAGCTTTGGCTCAATGCGGAGATGAGGTGTTGCGGCTGCTCATTTG 2196  
Qy 2149 AGCTTCATCTCTGCTTATGATGTGTTGCGGATCTGCGGCGGCGGCTGATGCTG 2208  
Db 2197 AGCTTGAACGTGTGCTTATGATGCTGCTTTCGCTTATGTTGAGGCTGAGCGCTGTA 2256  
Qy 2209 GTGCGGAGCGGCTCAAGCTGCGGATCCGCGCATTTGGGAGAGTTGATGAAAGAG 2268  
Db 2257 CTGCGGAGAGAGACGAGGAAAAAGATCGGCTCATTTGCGACATGATGACACAGT 2316  
Qy 2269 AAGGTACGATGTGAACTCGGTGCGGCGCTGATGCGGATGCTGTGACATTTGAG 2328  
Db 2317 CATGTAAACGCTGGAATGACGTGCGCGCATGATGACATGTGTGCG---AATACAC 2373  
Qy 2329 GGTGCGCGGATGCTGCTGATGATGCTGTGCGGCTTTGCTGCTGAGGCGGACGTGATC 2388  
Db 2374 AGCGCGATCGGATGATGATTCGACGTTGTGCTGTGACATGTTGAGGCGGACGTGATC 2433  
Qy 2389 CCGGTGCGGCTGCTGCGGCGA---GCTCCAGGCAATCAGGCGCGGCTGTGCTGATCAG 2445  
Db 2434 CCGCTTAACGTTAACGAGACATGCGGCGGCTCAATGAAAGATGACATCTACGT 2493  
Qy 2446 CTGCGCGGCGCACCGAGCGTCAATCTGTCTCATCTGAGTACCCGCTGAGAACTGAC 2505  
Db 2494 CTGCGTGAAGGACCGAGTGGCCATCTGTGCTACTACCCGATAGAGGTGAA 2553  
Qy 2506 CTATGTGCGGAGCATCTCTACGCGCTGCTGCGCAACGACCTTCAAGTGTCTC 2565  
Db 2554 TCGACGTGACCAAGATTTCCCTACGCTGCGGCGCTGCGCAACGACGATATCTGCTA 2613  
Qy 2566 GATGAGCGGCTGCAACCGCGCGGCTGTGAGTTCCGCGGCAACTATGATGCGGCGGCTC 2625  
Db 2614 AATGCGCACTGAGAGAAATGTCGCTGCGGCTGAGAGAGATTTGATTTGCGGAGTG 2673  
Qy 2626 GGGCTGCACTGGGCTACTGCGGAGTGAAGAGAACGCGCAAGAGCTTCTCTGTGAC 2685  
Db 2674 GGGCTGCGCAAGGCTACTGAACGACGACAGAAACGCGGAGAGCTTTGTCTGCGCG 2733  
Qy 2686 CCCGAGACCGGAGAGCGCTTACAGACCGGCGATCTGGCGCTGCTACCTGCGGATGA 2745  
Db 2734 GAAAGAGTGTGAGCGAATTTACCGCACTGCGGATTCGCGGCGCTACTTTGTGACGCG 2793



QY 2746 AACATGAGTTCAATGGGGGCTGAGCAACAAATCAAGCTTCGCGAATACCGCGTTGAG 2805  
 DB 2794 CAAGTCGCTTTTGGGGGCGCAACGATACCAAGTAAAGTGAATGGTTACCGTATTCGAA 2853  
 QY 2806 CTCGGGAAATCGAGGAAACGCTCAAGTGGCAATCCGAAGTACCGAGCCGGGATTTGAG 2865  
 DB 2854 CTGGGGAAATCGAGGCTGTCATTTGGCGCAATCCGATGTGAGCAATCAAGTGTGTG 2913  
 QY 2866 CCCGTGCGGAACGACCGCGCGAACAACTCTTCTAGCTATGTGTCCCGAGGCGACA 2925  
 DB 2914 GCAGTGGGTA-----TTCTGAACATCGTCGG 2940  
 QY 2926 CCGAGACGCGCTCCGAGACGACGAGCTTCAGACCGAGCGATTCACCGGAGACA 2985  
 DB 2941 CTGTGCTTTTTCCTCAACTGACGATCCGACAGCGCCAGGCAATTCGAAGTAAAGAA 3000  
 QY 2986 CAGCGCCGGAAGCGGACGCTTGAACGACGCGGAGAGGGTGCAGTTCAAGCTCGCTCA 3045  
 DB 3001 -----GCGGAGCGCGGACCTGCGCGAGGGATTAATGTGAATCCGCAACGCGTCA 3054  
 QY 3046 CACGACCTCCGAGGACCTGACGGAAGCCCGTCGATCTGACCGGCGAGATCCG 3105  
 DB 3055 GCGTTCAAACTAAGAGAGCAATATTCGGGCGTGGATGTCTGGGCAATGCACTGAG 3114  
 QY 3106 CCGGAGCGCGGCTGACCTTAACGCGCTCCGCTGAAGCTCCGAACGTTCTTGAGCC 3165  
 DB 3115 GCACCGCGGATAGACACGTTACATCAACCGCGGATATCGTATTCAGCCGCGCA 3174  
 QY 3166 CCGATTCGCTTTTGTGATTTGGTTCGATTCCTGAGCTGTGAGACCGGAGCCGAC 3225  
 DB 3175 AAAACCACTGACACGTTGGGCAATTTGCTGTGGGCGAGATCGCTTACCC 3234  
 QY 3226 GCGGCGACCTTCCCAATTCGTTATTCATCGCGCGGACGACGTAACCGGTCAAC 3285  
 DB 3235 GGTCTACCTTTTCCAAAGTATGCTATGGGTCCGCGGGGGCTATACCGGGCAAC 3294  
 QY 3286 TACGCTAATGTAATTCGCGCGCATCGAGGCGTGAAGAGGGCTTCTATTATTCAC 3345  
 DB 3295 TACGTGTACCTGTATCAGACAGATTCGAAGAGGGATATCCGATTTTACTACTTCGAC 3354  
 QY 3346 CCGTTGAGACGCTTTGCTGAAGCT---CTCCGATCAGCGGATTCAGAGCGCGAGCGAC 3402  
 DB 3355 CCGGACAGAGCTGTATTATCCGGTAGACACGAAGTGAAGCTGAACAGTGTTCAT 3414  
 QY 3403 GTTCGCAAACTTCGACGCTTTCGATGAAGCGCGCTTCAACCTCTGTTGCGGACAG 3462  
 DB 3415 GCCGACCTAATCAGTCTATTCGCGTAGACACGAAGTGAAGCTGAACAGTGTTCAT 3474  
 QY 3463 ATCGAGCATCGAGTGTGCTGTATG---ATGCTGTCCGAGAAATTTGCTGCTGAG 3519  
 DB 3475 ATGGCGGTATCTCGCATTTATAGGCAAGAGAGCTTGGACATCTTCGAGTAGGA 3534  
 QY 3520 GCCGATATATGCGGAGCTCTGATGAGACGAGCGCTTCTCGAACAATCGCGCTGT 3579  
 DB 3535 GCAAGTACTCTGCGCATTTACTGGAAGAAATGCGCGCTACCGATTTGGGGCTGTGC 3594  
 QY 3580 CCGGTGGGCAATTCATTTTGAACAGTTCGCGCGCTTCGACCTTCGCACTTCGAGC 3639  
 DB 3595 CAACCTGGAGTGGCAAGATTTTCCGCTGTGGCATTCGATTTCAATTTGCGCACATCAT 3654  
 QY 3640 GTTTACTGTCAGCGCATGCTGGGCGGGGGGTAGACCCGCGAGTTCCAGGCTCTAG 3699  
 DB 3655 CGCTATGTCCATTTGACCGGTGGGGGGCGGATAGGCAAGAGCGCAAGTGTGAGCA 3714  
 QY 3700 CTCGTGACGAGATTCCTACCGAGGCGCGCAACGCGCGCGCTCCCGCGCGAG 3759  
 DB 3715 TTGCTGCGGATTTCTTCACCTATGGAAGCAAGAAACCGCTGCGCGCTGAGCAG 3774  
 QY 3760 CAGACCTTGCAGATATGCTTCGCGACTTTTGAAGACCAAACTACCGCATCATGTG 3819  
 DB 3775 CAGAGCTAACAAAGATCGCATGCTGCGCGGCTGCGTACGAACTGCTGATATGTG 3834

QY 3820 CCTACAGTCTTGTGAGAGCTGATGCTGTCGCGTGAAGCTCCAGGAGGTGATGCT 3879  
 DB 3835 CCGAGTGTATGTATGTAGCCAGCGATTTTCCGTAAACCGCAAGCTAGATGCG 3894  
 QY 3880 AAGGCTTCGCGGACCGGAAGATACCTGTGCGCGCGCAT-----TCGGGGCAGACG 3933  
 DB 3895 CAAAATTTACAGCTCAGAGGCGCAACAATTTGCCACAGCGTGAAGGGGTGCTCAATC 3954  
 QY 3934 GCGGACCGGACGCTTGAAGAGATCTGTGCGGTGTAACGAGAGGTGCTCGGCTG 3993  
 DB 3955 CAGGTGACAGTGTCTTACCAACAGCGCGTGGTGGCTCTGCGACAGAGTACTCGGCTG 4014  
 QY 3994 GAGGTGTGCGGCTCCACAGAGCTTCGTCAATCTTGTGCGACATTCATTCATCTGT 4053  
 DB 4015 ACCCAGTGTGCGCGGAGACGATTTCTTGTGCGGGGCGAGTTATAGAAATTTGTTG 4074  
 QY 4054 CGCATGAGAGGCTGTTCAGAAAGAGGCTGATAGGAGATTCGATCAGGATTTGTC 4113  
 DB 4075 CGTATTCAGCAGGACCTTGAGGCGATTTATCGGCGAGAGATTCATTCATGCTGTT 4134  
 QY 4114 CAGTACCGGAACTCGGCTCGCTGCG 4140  
 DB 4135 CGTGTCCAAACCATCGCGAATGTGCG 4161  
 RESULT 5  
 ID ABA89188/c  
 AB ABA89188 standard; DNA; 48715 BP.  
 XX  
 AC ABA89188;  
 XX  
 DT 11-FEB-2002 (first entry)  
 XX  
 DE Escherichia coli polynucleotide SEQ ID NO 921.  
 XX  
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;  
 XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;  
 KW systemic infection; non-diarrhoeal infection; septicemia;  
 KW pyelonephritis; antibiotic resistance; ds.  
 XX  
 OS Escherichia coli.  
 XX  
 PN W0200166572-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 12-MAR-2001; 2001WO-EP03445.  
 XX  
 PR 10-MAR-2000; 2000FR-0003145.  
 XX  
 PR 02-FEB-2001; 2001FR-0001449.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Bingen E, Bonaccorsi S, Clermont O, Naessle X, Tinsley C;  
 XX  
 DR WPI; 2001-550253/61.  
 XX  
 PT A library of DNA fragments of Escherichia coli strains for the  
 PT phylogenetic determination of a given strain comprises polynucleotides of  
 PT nature B2/D+ A- -  
 XX  
 PS Example 6; Fig 6; 646pp; English.  
 CC  
 CC The invention relates to a library of DNA fragments of Escherichia coli  
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)  
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature  
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,  
 CC antibacterial and immunosuppressive activity as part of pharmaceutical  
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli  
 CC infections. The polypeptides are useful for determining the phylogenetic  
 CC group of a given E. coli strain. These polypeptides can detect and treat  
 CC an undesired development of E. coli, particularly an extra-intestinal  
 CC infection that include systemic and non-diarrhoeal infections such as

CC Septicemia, pyelonephritis and meningitis this is particularly  
CC advantageous as bacterial resistance is increasing with the more  
CC frequent use of broad spectrum antibiotics.

XX Sequence 48715 BP; 11934 A; 13488 C; 12203 G; 11090 T; 0 other;

Query Match 12.8%; Score 543.4; DB 22; Length 48715;  
Best Local Similarity 48.5%; Pred. No. 8.5e-106;  
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;

```
Oy 181 ATCCGCCCCCGCCGAGGCGGACGCTCCGTTTCTCTACAGACATCCAAATCC 240
Db 18373 ATAGGACAGACCCCTGAGCATCTGTCAACCGTTTCCATTGACGAATGTTCAAGCCGC 18314
Oy 241 TACTGCGCGGCGGACAGAGCGTTTACGGTCCCGAGGGATCCAGCCCTATCGCGA 300
Db 18313 TACTGCGTTGAGCG--ACAGAGGGGTGCTACTGTGATCGGACCCAACTTACCATGAA 18257
Oy 301 TACGACTTACGAGATCTGACGCTGCCGAGGCTGAGCCGCGCTTTCGAAAGTCTGCG 360
Db 18256 TTTGACGTAAACAATTAAATGTAAGCGTTTACCCATGCGGTGAATGCGCTGATCCGT 18197
Oy 361 CGGACGACATCTCTCGGCGCCCAACGCTGCCCGGACATGATGACGTATGAGCTTAA 420
Db 18196 CGCATGAAATGCTACGTGCGCGGTACTCCCGACGGTACTCAGACAGATTCCTGCGCA 18137
Oy 421 GTCCAGCG--CGACATCGAATCATGCTGCGCGGGCTGACCGGAGCACAGGGA 477
Db 18136 GTGCGGCGGTACATTAAGAGAGCGCGATCTGATGCTTGTCTCCCTTACGCAAGAAAC 18077
Oy 478 GCGAGGCTGTGTCTGTCGAGATGAGTGTGCAACCGCATCTATGACACCGAGCGCCT 537
Db 18076 GATGCTTGATGAGCATCCGCGATCGGCTGTGTCATGTCATCCGCGAGATGTTGG 18017
Oy 538 CCGCTCTACATCTGCTGCGCGCTGCGTGAACAGCGGCAAAACCGTCTGCTCACT 597
Db 18016 CCGCTTTGATTGATTAGTTATTCGCTTGACAGCGGCAACATGCGCTTGCAATTTCAGT 17957
Oy 598 ATGCATCTCATTAAGTTAGCTTACCTAGGACGCTGTCCATCATCTTCAAGACTGCTCAGC 657
Db 17956 CTGCATCTGCTGATTCGCGATCTCTGATGATGCGGACGCTACAGAGAGTGTATGATG 17897
Oy 658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGAGCTCTCTACCGCGATTATGTA 717
Db 17896 CTGTACCGTGAAGCCCATGTGTCTACCTGCGTGTCTGTAACGTTCTCTTTCGTGATCAG 17837
Oy 718 CTCGCGCTGAGCTCTCGCAAGATCTGAGCGCATCAACGATGATGATTACTGGAAG 777
Db 17836 CAGGCGCTGTGTGTAGAGCAGGCGGAGTGAAGCTTATGACGCGATCAGGCGCTATTGGCA 17777
Oy 778 CGGCGCATCGCGAGCTCCACCTCGCGGAGCGCTTCGATGAGGCGGATCCATCTAC 837
Db 17776 CGGCGCTGCGGACCTGTATGCGCCACCAACGCTGCGGTCAGAGGCGATTGGCGCA 17717
Oy 838 CTGAAGAGATCCGCTTCGAGCACAAGAGCAATGCTGCGCTGCGACTCTCTGGGATGCA 897
Db 17716 CTGTCTGGATGAGTTGTGATGCTGCGGCTATCGGCTGTACGCCCAACCTGGGAGATG 17657
Oy 898 TTGAAGCGCGTGTGCGGAGCGCGGCTGACCCCGAGCGGCGTCACTCTGCTGCTCATTT 957
Db 17656 CTGAGCGGCTGTGCGCAACGCAACGATATCACCAAGACGGCAATTTGTGTGACATCTTT 17597
Oy 958 TCCGAGGATCGGCGCTGAGCGCGAGCGCCCGGTTTACGCTCAACATAAGCTCTTC 1017
Db 17596 AGCCAAAGTGTGCGACGTTGAGCCTTACCGCGAGCTTACGCTCAATCTGACGTTGTC 17537
Oy 1018 AACCGGCTCCCGTTCATCCGCGCGTGAACGATATCACCGGGGACTTCAAGTCAATGATC 1077
Db 17536 AACCGCGCGAGGTTACCCCAAGCAGAGCATTTATGATGATTTTACCGCTGTACG 17477
Oy 1078 CTCCTGACATGACACCACTGCGACAAAGACTTGAACAGCGGCTGTAAGCTGATCA 1137
Db 17476 TTGCTGAATGTTTGTAAAGACAGCAGCACTCTTATGCGCCACAAAGCTCAGGATTCAG 17417
```

```
Oy 1138 GAGCAGCTGTGGGAGAGCATGATCATGCGACGTAAACGGTAT-----CGAGTTC 1188
Db 17416 GTGCAACTGTGGAAAGATCTGAAACATGTGTGTTTCAAGTGGGATCCGCGCAAGCGAGCG 17357
Oy 1189 CAGCGAGAGCGCGCGCGGCTCTGGGATCCAAAGAGCGCATTTGTTCCCGTGTCTC 1248
Db 17356 CTGATCCATTAACCGGTCTGTTTCAATGCGCGATGCGGTGTGTATCATGATGTTGAT 17297
Oy 1249 ACGAGCGGCTTTAACAGCAAGTGTGTGTCACTCTGTTGACAGAGCTGGAACTCG 1308
Db 17296 ATCGAGGGAGAGAGACTGCGCAAGACCTTGGGACAAACCGTTTATCTGTGTCCG 17237
Oy 1309 GTGTACACAGACAGCAAGACTCTCAGCTGCTGTGATCATCACTCAAGCTTACAGACCAT 1368
Db 17236 GAGCGCAATATTATCCCAAAACCGAGGTGTGTGTGATCATCAAGTATGATGATGATGCT 17177
Oy 1369 GGGGACTCTGCTCGCGTGGACATGTCGACGAGTGTTCGCGCCGACCTTGTGAC 1428
Db 17176 GGGGAGTTGCAATTCACTGGGAGCGCGGTGAGCAACTGTTGATACCAAGCTCTGGAT 17117
Oy 1429 GACATGCTCGAAGGTACGTGTTTCTCCGCGGCTCACTGAGGAACATGGGATGA 1488
Db 17116 CAGATGTTGATGCTTATTTGATGCTGCTGACGAGCGCTGTGTCATGCC-----G 17066
Oy 1489 CAGGTGCGCTGTTCCGCTTCCGCTGCGCAGCTTAAAGCGCGGAGCGCAACGCGAC 1548
Db 17065 CAAGTGTGTGGGGGTAAATGATTTCTGTGCGCTGCCACCGTTTATGTCACCGCTACG 17006
Oy 1549 AACGCGCTGTAGAGAGCATACGTCGACGCGCTGTTTCGCGCGCGGCTGACGACTG 1608
Db 17005 CAGGCTCTGCACTACGCGCTTGTGTCACCATGATTAACGCTCAGGAGCACTGACG 16946
Oy 1609 CCATGACAGCTCGCGTGTGTGCGCGGCGGCAAGACGCTCAAGTCAAGAGCTTTCGGGC 1668
Db 16945 CCACAGAACTGCGCTGATGATCTATCCGTGAATGACCTATGCCAATGTGACG 16886
Oy 1669 GTTTCGCGCAGCTTGTGCGCGCGCTGCGCAGAGAGGGGCGACGCGCCAAACATTTGTC 1728
Db 16885 GCGGCGATCATGTGTGCGCGCGCTGTTTACGCGGTGCGGTGAGCATGCGGACCGCGTG 16826
Oy 1729 GCGGTGTGATGAGAAAGCTGGAGACAGTTGTCCGCGTTCTGCGGTGCTGATGTA 1788
Db 16825 GCGGTGTGATGAGAAAGCTGGAGACAGATTGCGCGCGTACAGCGGATTTTACAGCTG 16766
Oy 1789 GCGCGGCTACGTCGCGCATGGAATCCGACCTACCGCGGAGCGATACCTACTCTC 1848
Db 16765 GGTGCGCTATCTGCGCATGGAATCCGGTGCTACCGCACAGCGTCCAGCTTTTGTG 16706
Oy 1849 GATCATGATGATGAAGCTGTGCTGACGAGCATGCTGATGCAAACTGTCATG 1908
Db 16705 ACGGTGCGCGAGGTGCGGTACAAAGTACGAGACCGGCTCTCAAGCAATTGAGCCGTG 16646
Oy 1909 CCGCGGAGATCCAGCGGCTGTCGTGAGCGAGCGCGGCTGGAAGGCGCGACCGACAG 1968
Db 16645 CTGCGCG-----TCTGATCATGACGAGGAATGTGGAACCGCTGCTCG 16598
Oy 1969 CTTCCGATGATGCCATTTCAACACCTTTCGATCTCGGTATGTCATTTACACCTCGGGA 2028
Db 16597 CCGTTGCTGAAGTGTGCGGAGATGTCAGGATCTGCGCTATATCATTTTACCTTCGAG 16538
Oy 2029 TCCAGAGGTTGCCCAAGGGGATGATGATCATCATCGGGGTCCGTCACACCATCTCTG 2088
Db 16537 TCCACCGTACCCGAAAGAGATGATATCAACACCTGCGGCGCATTAACAGCTGGAA 16478
Oy 2149 AGCTTCATCTCTGCTTATGATGTGTGCGGATCTGCGCGGCGGCGGTATGATCTGTG 2208
Db 16417 AGCTTGAACGTGTGTTTGAAGATGCTTTCGCGCTTTTATGATGAGGTGAGCGCTGTGA 16358
```

QY 2209 GTGCCGAGACCGCTCAAGCTGGCCGATTCGGCCGATTTGGGCGAGATTGATCCGACGAG 2268  
 DB 16357 CTGCCGGAAGCAGAGCGGAAAAAGATCCGCTGATTTGGCGACAGATTATGGACACGAT 16298  
 QY 2269 AAGGTGACGGGTGGAATCTCGGTCCGGCGCTGATCGGATGCTCTCGACATTTTGA 2328  
 DB 16297 CATGTAAAGGTCTGGAATGCAAGTCCCGACCTGATGATGCTGTGCG--AAATCCAC 16241  
 QY 2329 GGTCCGCCGATTCGCTCGTGAAGTCTTCCGCGCTTTCCCTGCTGAAGCGCGCATGTGATC 2388  
 DB 16240 ACGCGGATCGGATGATGATTATCCGACGTTTCCGCTGAGCACTGTGAGCCGCGCATGATC 16181  
 QY 2389 CCGGTGGGCTGCTGCGCA---GTTCCAGGCCATGAGGCCGCGCTGCTGATGATCAGC 2445  
 DB 16180 CCGCTAACGTTACCGGAGCAGATGCGGAGCGGCTCAATGAAACGATGACATCATCAT 16121  
 QY 2446 CTGGGCGGGCCACCGAAGCGTGCATCTGCTCCATCGGGTACCCCTGAGAACTGCAC 2505  
 DB 16120 CTGGGTGAGCAGACCGAGTCCGCTATCTGCTGCTTACTACCCGATAGGTAGTGGA 16061  
 QY 2506 CTATCGTGGGCGAGCATCCCTTAAGCCGCTGCTGCGCAACGATTCACGTTCTC 2565  
 DB 16060 TCGACGTGAGACGATATCCCTACGCTGCGGCGCTGCGCAACGACGACGATAGTGTCTA 16001  
 QY 2566 GATGAGCGCTGACACCGCGCCGCTGCTGCTGCTGCGCAACGATTCACGTTCTC 2625  
 DB 16000 AATGCCAACCTGAGGAAATGTCGCTGCGGCTGGAAGAGATTGCAATTCGCGGAGT 15941  
 QY 2626 GGGCTGGCACTGGGCTACTGCGCGATGAAAGAGAGAGCGCAAGCTTCTCTGTCAC 2685  
 DB 15940 GGGCTGGCAAGGCTTACTGCGCGATGAAAGAGAGAGAGCGCGCGAGCTTCTCTGTCAC 15881  
 QY 2686 CCGGAGACCGGGGAGCGCTCTAACAAGACCGCGATCTGGGCGCTACCTGCGCATGGA 2745  
 DB 15880 GAAGCAGATGATGAGGAAATTTACCGCATGCGGAGATCGCGGCGCTACTTGTGTCAGCG 15821  
 QY 2746 AACATGAGTTCATGCGGCGTGAAGCAACCAATGAGCTTCCGCGATACCGCGTGA 2805  
 DB 15820 CAAGTGCCTTTTGGGCGCAACGATCCCAAGTGAAGTGAATGATGATGATGATG 15761  
 QY 2806 CTGGGGAATCGAGGAACGCTCAAGTGCATCCGACGTCAGCGACGCGGATGTTGT 2865  
 DB 15760 CTGGGGAATCGAGGCTGATTTGGCGCATCCGACGTCAGCGATGATGATGATG 15701  
 QY 2866 CCGGTGGAACGACCGCGGAAACAAGCTCTTCTGACCTATGCTGCTGCGGAGCA 2925  
 DB 15700 GCAGTGGTAA-----TTCTCAACATCTGCTG 15674  
 QY 2926 CCGGAGCGGCTGCGGAGCAGACGCGGCTCAAGACGAGCGGATTCGACGAGAGCA 2985  
 DB 15673 CTGGTGCCTTTTCCAACTGACGATCCGACGAGCGGATTCGACGAGAGCA 15614  
 QY 2986 CACGCGCGGAACGAGCGCTTGAAGCAGCGGAGGAGTTCAGTTCAGCTCGCTGCA 3045  
 DB 15613 -----GCGGAGCGGCGGACCTGCGGAGGATTAATTTGAAATCCGAGCAGCGCTGA 15560  
 QY 3046 CACGAGCTCCGAGGAGCTGAGCGGAAAGCCGCTGCTGATTCGACCGGCGAGATCCG 3105  
 DB 15559 GCGTTCAACTCAAGAGGACCATATTCGCGCGCTGATGCTGCGCATTTGACGATGACG 15500  
 QY 3106 CCGGAGCGGCGGCTGAGCGTCTACGCGCGCTGCGGATGCGGAAAGTTCTTGAAGCC 3165  
 DB 15499 GCAACCGCGATGACACGTTACATCAAGCGCGAGCTATGCTCATTTACGCGGCA 15440  
 QY 3166 CCGATTCCGTTTGTGATGCTGATTCCTGAGCTGTTGAGCAGCGTGAAGCCGAC 3225  
 DB 15439 AAAACACGCTGACACGTTGGGCAATTCCTGCGGCTTGGGCGAGATGCGTCAAC 15380  
 QY 3226 GCGCGACCTTCCCAATTCCTTATCATGCGCGGCGACGATACCGGTGCAAC 3285  
 DB 15379 GGTCTACTTTTCCAAAGTATGCTTATGCTCCGCGCGGCGGCTATACCGGTGCAAC 15320  
 QY 3286 TAGCGGATGTAATCCGCGCGATCGAGGCGTGAAGAGGCTTCTATTTACAC 3345

DB 15319 TAGGTATCTGATCCAGACAGATCGAAGAGGAGATCCGATTTACTACTTCCAC 15260  
 QY 3346 CCGTTGAGCACCGTTTCTGAAGCT---CTCCATCAACGAGATGACCGCGAGCGAC 3402  
 DB 15259 CCGGACAGAGGTGCTTATGCGGATACACAGAACTGAGCTGAACGATGTTTCAT 15200  
 QY 3403 GTTCGCAAACTTGCAGTGTTCATGAAAGGGGTTCAACCTCTCTGCTGGGCAAG 3462  
 DB 15199 GCGGACCTTAACAGTCAATGCGATCGGGGCGATTCAAGCTTTATGCTGAT 15140  
 QY 3463 ATGACGCGATGATGCTGATG---ATGCTGCTGCGGAAATTTGCTGCTGAG 3519  
 DB 15139 ATGCGGATGATGCTGATGCTTATGAGGAGAGGAGCTTGGCACTTCTGCTGATGAA 15080  
 QY 3520 GCGGATATATGAGCGACCTCTGATGAGAGAGCGCTTCTGCAACATCGGCTGCT 3579  
 DB 15079 GCAAGTACTCTGCTATTTACTGAAAGAGATGCGCGCTACGATGAGGCGCTGTC 15020  
 QY 3580 CCGGTGGGCAATTTCAATTTGAACAGTTCGCGCGCTTCTGACCTGCGATTCGAC 3639  
 DB 15019 CAACTTGGATGAGCACTTTCGCTGCGGATGCAATTTCAATTCGCGCATCAT 14960  
 QY 3640 GTTACGTCAGCGATGCTGAGCGGCGGATGAGACCGCGCACTTCAGGCTGTACG 3699  
 DB 14959 CCGTATGCTATTCATGACCGTGGGCGCGATGAGGAGGCGGCAAGTCTGACAG 14900  
 QY 3700 CTGGTCAAGATTCCTCAACGAGCGCGCACGAGCGCGCGCTTCCGCGCGAG 3759  
 DB 14899 TTGCTGCGGATTTCTCCACCTATGAGAAACGAGAGAAACCGCTGCGCGCTGAGCATG 14840  
 QY 3760 CAGCACTTGCAGATATGCTTGGCACTTCTTGAAGCAACCACTCCGAGTACATG 3819  
 DB 14839 CAGACTCAAAAGATGCACTGCTGCGCGCTGCTGAGCACTCTGACTATATG 14780  
 QY 3820 CTTACAGTCTTGTGAGAGCTGATGCTGCGCTGAGCTGCAACGAGGAGTCTGATG 3879  
 DB 14779 CCGAGTATGATATGATGAGGAGCGATTTCCGTTAACCGCTAACGAGCATGATG 14720  
 QY 3880 AAGCCCTGCGGAGCGGAGATGATCTGCTGCGCGCAT-----TCGGGCGACAG 3933  
 DB 14719 CAAAATTTACACTCAGGCGCAACAAATGCGCACAGCTGAGCGCGTGGTCCATC 14660  
 QY 3934 GCGGACGAGGAGCTTGGAGAGATGCTGCGGCTGCTGATGAGGAGTCTGCGGCG 3993  
 DB 14659 CAGGTGACAGTGCCTTACACAGCGGCTGCGGCTTGGCAGAGATCTCGCGTG 14600  
 QY 3994 GAGTGTGCGGCTCAGCAGAGCTTCTGATCTTGTGCGACATTCATTCATCTG 4053  
 DB 14599 AGCAGCTGTGCGCGGAGAGCATTTCTGCTGCGGCGGAGTCTATAGATTTG 14540  
 QY 4054 GCGATGAGAGCTGTTGCAAGAGAGCTGATGAGGAGATGCGATCAACGATTTG 4113  
 DB 14539 CGTATTCAGCAGGACTTGAAGCGATTTACGCGGAGAGATTCATGCTGATCTG 14480  
 QY 4114 CAGTACCGGACCTGCGCTGCGGCG 4140  
 DB 14479 CGTCTGCCAACATGCGGATGTTGGG 14453

RESULT 6  
 ABST8890/c  
 ID ABST8890 standard, DNA; 50538 BP.  
 XX  
 AC ABST8890;  
 XX  
 DT 17-DEC-2002 (first entry)  
 XX  
 DE E. coli CF703 genomic sequence #57.  
 XX Pathogenic; Escherichia coli CF703 infection; livestock; pyelonephritis;  
 KW urinary tract infection; open reading frame; ORF; uropathogenic;  
 KW antibacterial; aropathic; nephrotropic; gene; de.

XX Escherichia coli.  
OS  
XX  
PN W0200259320-A2.  
XX  
XX 01-AUG-2002.  
XX  
XX 19-OCT-2001; 2001WO-US46833.  
XX  
XX 19-OCT-2000; 2000US-242412P.  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Blattner FR, Welch RA, Burland VD;  
XX  
XX WPI, 2002-691532/74.  
XX  
XX New DNA sequences of the pathogenic Escherichia coli CFT073 strain,  
PT useful for preventing or treating E. coli CFT073 infection in humans or  
PT livestock -  
PS  
PS Claim 1; Page 186-214; 765pp; English.  
XX  
XX The present invention relates to polynucleotide sequences from the  
CC genome of the pathogenic Escherichia coli strain CFT073. Almost all  
CC the sequences present in E. coli CFT073 are absent in the previously  
CC sequenced laboratory strain K-12. The polynucleotide sequences of  
CC the invention are useful for preventing, diagnosing or treating  
CC E. coli CFT073 infection in humans or livestock. The polynucleotide  
CC sequences are useful for preventing urinary tract infections and  
CC pyelonephritis. Likewise, the polypeptides encoded by the different  
CC open reading frames (ORF1-5) are useful for generating a vaccine  
CC against uropathogenic E. coli strains. ABS7834-ABS79085 represent  
CC genomic sequences from E. coli strain CFT073.  
XX  
XX Sequence 50538 BP; 12312 A; 13865 C; 12701 G; 11654 T; 6 other;  
SQ  
Query Match 12.8%; Score 543.4; DB 24; Length 50538;  
Best Local Similarity 48.5%; Pred. No. 8.5e-106;  
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;  
QY 181 ATCTGCCCCCGCCGACCGGAGCGGACGCTCCGTTCTCTCAGACATCCAGAAATCC 240  
DB 20201 ATAGGACAGACCCCTGAGCATGTCATCAACCGTTCCCATTAAGCGATGTTACGGCC 20142  
QY 241 TACTGGCTGGCGGCGGACAGAGCGTTTACGTTCCCAAGGGATACAGCCCTATGCCAA 300  
DB 20141 TACTGGCTTGGACG--ACAGACGGGCTCTACCTGATCGGACCCACATCTAACATGAA 20085  
QY 301 TACGACTGTACGATCTCGACGTCGAGGCTGAGCCGCGCTTTCGAAAGTCTGCG 360  
DB 20084 TTTGACGTAGAACATCTTAATTTACGCGTTTACCCATGCGGTAAATCCGTATCCGT 20025  
QY 361 CGGACGACATGCTTCCGCGCCACACGCTGCCGACATGATGACAGTATGAGCCTTAA 420  
DB 20024 CGCCATGAATGCTACGTGCGGGTACTCCCGACGATACAGAGATTTGCGGAA 19945  
QY 421 GTCCAGCG--CGAATGGAATCATCATCTGCGCGGGCTGAGCCGAGACACAGGAA 477  
DB 19964 GTGCCGGTATCACTTAGAGAGCGCATGATGCTTTTCCCTTAAACGCAAGAAAC 19905  
QY 478 GCGAGGCTCTGTCTGTCGAGATCGAGTGTGACACCGCATCTATGACACCGAGCCCT 537  
DB 19904 GATGCTTGAATGCGATCCGCGATCGGCTGTGCAATCATGTGCAATCCGCGAATGTTGG 19845  
QY 538 CCGCTTATCACTGCTGCGCGCTGCGCTGACGAGCGGCAAAACCGTCTCTGCTCACT 597  
DB 19844 CCGCTTATCACTGCTGCGCGCTGCGCTGACGAGCGGCAAAACCGCTTGAATTTCACT 19785  
QY 598 ATGATCTCATTAAGTTGACCTAGGACCTGTCTCATCATCTTCAAGACTGCTGCTACG 657  
DB 19784 CTCGATCTGCTGATGCGGATGCTGTGATGATGCGCACGCTAACAGAGAGTTGATGATG 19725

QY 658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCTGAGCTCTGCTACCGGATTAATGTA 717  
DB 19724 CTGTACCGTGAAGCCCATATGTGTCACTGCGTTGTGCTACCGTTCTCTTTTGTGACTACGG 19665  
QY 718 CTCGCGCTGAGAGTCTCGCAAGAGTCTGAGCGCATCAACGATCGATGATTACTGGAAG 777  
DB 19664 CAGCGCTGTTGTGATGAGACAGCGGATGAACTATGACCGCATCAGGCCCTATTGGCAA 19605  
QY 778 CGGCGCATCGCGGACCTCCACCTCCGCGAGCGCTTCCGATGAAGCGGATCATCTACC 837  
DB 19604 CGGCGCTGCGGACCTGTATAGCCCATCAACGCTGCGGTATCAGGCGGATTTGGCGAA 19545  
QY 838 CTGAAGAGATCCGCTTCCGACACAGAGGATGCTCCGCTGCACTCTCGAGCTGCA 897  
DB 19544 CTGTCTGCGATCAGGTTCTGTACGTGCGGCTATGCGCTGTACGCCACACCTGGGAGTG 19485  
QY 898 TTGAACCGCGCTGTGCGGAGCGCGGCTGACCCCGACGCGGCTATCTGCTCATTT 957  
DB 19484 CTGACCGGCTGGGCCCAACGACACATATCAACAGAGGCAATTTGTTGACAGCTCTT 19425  
QY 958 TCCGAGTATCGGCGCGCTGAGCGCGGACCGCCCGTTTACGCTCAATTAAGCTCTTC 1017  
DB 19424 AGCCAGTGTGCGACAGTTGAGCCTTAGCCGACGTTTACGCTCAATGACGTTGTC 19365  
QY 1018 AACCGGCTCCCGTTCATCCGCGGTGAACGATATCACCGGAGCTTACGCTCATGTC 1077  
DB 19364 AACCGCGCGAGGGTTACCCCAAGCAGAGGAGTATTTGATTTTACCGCTGTACG 19305  
QY 1078 CTCTGACATGACACCACTTCGCGACAGAGCTTTCAGACGCGCTTAAGGCTATTCAA 1137  
DB 19304 TTGCTGAATGTTTGTGAAGACAGCGACACTTATGCGCCACAGCTCAGGATTTGACG 19245  
QY 1138 GAGCAGTGTGGAAGAGATGATGATCTGCGACGTAACGCTAT-----CGAGGTC 1188  
DB 19244 GTGCACTGTGGAAGATCTGAAATCTGTCTTCACTGAGATCCGCCACGAGGCG 19185  
QY 1189 CAGCGAGAGCGCGCGCGCTGCGGATTCACAGCGGATTTTCCCGTGTGTC 1248  
DB 19184 CTGATTCATACCGGCTGTTTCATATGCGCGGATTCATGATGATTTGAT 19125  
QY 1249 ACGAGCGCTTAAACAGCAAGTGTGTCTGCTGATCTGATCACTGAGTACAGTACAT 1368  
DB 19124 ATCGACGGGAGAGACATCGCCAGAACCCCTGGGACACAAACCGTTTACTGTGTGTCG 19065  
QY 1369 GTTAAACACAGACAGACATCTCAGCTGTGCTGATCTGATCACTGAGTACAGTACAT 1368  
DB 19064 GACGCCATATTATCCCAACACCGAGGTGTGCTGATCACTGAGTATGAGTGTGCT 19005  
QY 1369 GGGGACCTGTCTCTGCGCTGGAGACATCTGACAGGAGTGTCCGCGGACCTTCTGAGC 1428  
DB 19004 GGGGAGTTGATTTCAATGAGACGCGGTGAGCAACTGTTTGAATACACGCTGTGAGAT 18945  
QY 1429 GACATGCTGAAGCGTACGTGTTTTTCTCCGCGGCTCACTGAGAAACATGGGTTAA 1488  
DB 18944 CAGATGTTGTGCTTATTTATGATGCGCTGACAGGCGCTGTGCCATCC-----G 18894  
QY 1489 CAGGTGCGCTTGTGCTTCCCGCTGCCAGCTGAAGCGCGGAGGCGAAACGCGAAC 1548  
DB 18893 CAAGTTGTGTGGGGTAAATATGTTCTGTGCGCTGCCACCGTTATGTCACCGGTACG 18934  
QY 1549 AACGCGCTGTGAGGACATATCGTACAGGCGCTGTTGCGGCGCGGCTGAGCAGCTG 1608  
DB 18893 CAGGCTCTGACATGAGCGCTTGTGTGACCATGATTAATGCGTACAGGACATGACG 18774  
QY 1609 CCATGACGCTGCGCTGTGTGCGCGGCAAGACGCTCACGTAAGAAAGCTTTGCGGC 1668  
DB 18773 CCACAGGAACTGCGCTGATAGTCTTATCCGTGAATGACCTTACGCAACTGTCGACG 18714  
QY 1669 CGTTGCGGCGGACCTTTGGCGCGGCTGCGAGACGCGGCGGACGCGCAACATGCTG 1728  
DB 18713 GCGGCGATCATGTGCGCGCGCTGTTTACGCTGCGGCGTGCACATGCGACCGCGTG 18654  
QY 1729 GCGGTGTGATGAGAAAGCTGGGAGCAGGTTGTCCGCGTCTTCGCGGCTCGAGTCA 1788

Db 18653 GCGGTGTATGAAAAGGCTGGCAGAGATTGCCGCTACCGCATTTTACGACTG 18594  
Qy 1789 GCGCGGCGCTACGTCGATGCTGCGGACTACCGGCGGAGCGTATCCACTACTCTC 1848  
Db 18593 GGTGGCGTATTACTGTCAGTGGATCCGGTGTCTACCGCCACAGGTGCGCAGCTTTTGCTG 18534  
Qy 1849 GATCATGTGAGTAAAGCTCGTGTGACGAGCCATGCTGGATGGCAATCTGTATGG 1908  
Db 18533 ACGGTGGCGAGGTGGGCTGACAACTACGAGCCGGGTCTCACGCAATTTGAGCCGTG 18474  
Qy 1909 CCGCGCGGGATCCAGCGGCTGCTCGGAGAGGCGGCGGTGGAAGCGACGCGACAG 1968  
Db 18473 CTGCCG-----TGTGATCATGACGAGGAATGCTGACACGCTGCTGG 18426  
Qy 1969 CCTCCGATGATGCCATTCAGACACCTTGAGATCTGCGTATGCTACACTTCGGGA 2028  
Db 18425 CCGTTGCCGAGAGTGGCTGGGAGTGTACAGGATCTGGCTATATCTTTCACTTCGGC 18366  
Qy 2029 TCCACAGGTTGCCAAGGGGTGATGATGATCATGCGGGTCCGTCAACACATCTGG 2088  
Db 18365 TCCACCGGTACCCCGAAGAGATGATGACACACGTGCGGCTATGAACACGCTGGA 18306  
Qy 2089 GACATCAAGAGCGCTTCGAAATAGGGCCCGAGACAGGGTGTGGCGCTCTCGCTG 2148  
Db 18305 GACATCAAGAGCGCTTTGAGCTCAATGCGAGATAGGTGTTCCGGCTGTATATTTG 18246  
Qy 2149 AGCTTCGATCTCTCGTCTATGATGTGTTCGGGATCTGGCGGCGGCTACGATCGT 2208  
Db 18245 AGCTTTGACTGTGCGTTTACATGCTTTCGCTTTATGTTAGTGGGTGAGGCTGGTA 18186  
Qy 2209 GTGCCGAGCGCTCCAGCTGCGCGCATCCGGCCGCTTGGCGAGATTGATCGAACAGAG 2268  
Db 18185 CTGCCGAGAGCAGACGAGGAAAAAGATCCGCGTCACTTGGCAGACATTTAGCACACGT 18126  
Qy 2269 AAGGTAGCGGTGGAACCTCGGCGGCGGCTGATGCGGATGCTGTCAGACATTTTGG 2328  
Db 18125 CATGTAGCGTCTGGAGATGAGTCCGACATGATCAGATGCTGTGCG---AATACAC 18069  
Qy 2329 GGTGCGCCCGATTCGCTCGCTAGGTCTTCGCGCTTTCGCTGCTGAGCGGCGCATGATC 2388  
Db 18068 AAGGGGATCGAGTGAATTATCGACGTTTCCGTCTGCGCATGTTGAGCGGCGCATGATC 18009  
Qy 2389 CCGGTGGCGCTGCTGCGCA---GCTCCAGGCCATCAGGCCGCGGTGCGGTATCAGC 2445  
Db 18008 CCGCTAACGTATCCGAGACATGTCGAGCGGCTCAATGAACATGACATCATCAAT 17949  
Qy 2446 CTGGGCGGGGCGCACGAGGCTGATGCTGCTCAGGGTACCCGTAAGAGAACTGTCAG 2505  
Db 17948 CTGGGTGAGCGACGACGATGCGCATCTGTGCTACTACCCGATAGGTGAGTGA 17889  
Qy 2506 CTATCTGCGGCGAGCATCCCTTACGCGCGCTCGCTCGCAACGACGTTTCACTGCTC 2565  
Db 17888 TCGAGCTGACAGTATTCCTTACGCGTGGGCGCTCGCAACAGCAAGTATCGTCTA 17829  
Qy 2566 GATGAGCGCTGACACCGCGCGCTGTGCTGCTGCGGCGCACTTACATTTGCGGGGCTC 2625  
Db 17828 AATGCCAACCTGAGAGATGTCTGCTGCGGTGAGAGAGATTTGCAATTTGCGGGGAGT 17769  
Qy 2626 GGGCTGGCACTGGGCTACTGCGCGCATGAGAGAGACGCGCAAGGCTTCTCTGTCAGC 2685  
Db 17768 GGGCTGGCAAGGCTTACTGAGACGACGAGAGAAAAAGCGCGGAGCTTTGTCTGGCGC 17709  
Qy 2686 CCGGAGACCGGGAGCGCTCTTACAAAGACCGCGGATCTGGGCGCTTACCTGCCATGGA 2745  
Db 17708 GAAGCAGTGTGAGGAAATTATCCGACTGGGAGTGGCGGCGCTTACTTGTGACGGG 17649  
Qy 2746 AACATGAGTTCACTGGGCGTGAAGACAAATCAAGCTTGGCGGATACCGCGTTGAG 2805  
Db 17648 CAAGTGCCTTTTGGGGCGCAACGATACCAAGTGAAGTGAATGTTACCGTATCGAA 17589  
Qy 2806 CTGGGGAATGAGGAAACGCTCAAGTGCATCCGAAGTACGCGCGGCGGTATTTG 2865

Db 17588 CTGGGGAATGAGGCGCTGCAATTGGCGACATCCCGATGTGAGACATGATGGTGGT 17529  
Qy 2866 CCGTGGGAGACGACGCGCGCAAGACTCTTTAGCTTATGTGTCCCGAGGGCACA 2925  
Db 17528 GCAGTGGTAA-----TTCTCAATGTGTGCG 17502  
Qy 2926 CCGAAGCGGCTCGGAGAGGACGAGGCTCAAGACCGAGCGGATGACGCGAGAGA 2985  
Db 17501 CTGTGCTTTTGGCAAACTGACAGATGCGACAGGCGGAGGCACTTGCAACTTAAGAA 17442  
Qy 2986 CAGCGCGCGAAGCGGAGCTTTGAGCGACGCGGAGAGGTGCACTTCAAGCTCGCTGA 3045  
Db 17441 -----GCGAGGCGGCGGCACTGGCGAGGATATTATTGTGAATCCGCAACAGGCTTA 17388  
Qy 3046 CAGGACTTCGAGGAGACCTGGAAGGAAAGCCGTGTGATCTGAACGCGGAGATCCG 3105  
Db 17387 GCGTCAAACTCAAGAGACACATATTGCGCGCTGAGATGTGCTGGGATTCACATGAGC 17328  
Qy 3106 CCGAGGCGGGGCTGAGAGCTTACCGCGGCTGCGGATGCGGAAGCTTCCTGAGGCC 3165  
Db 17327 GCACCGCGGATGACACACGTTTACATCAAAGCGCGAGTATGTCATTTTCAGCGCA 17268  
Qy 3166 CCGATTCCGTTTGTGATTTGTGATTTGCTGATTCCTGAGCTGTTGAGCAGCGTGAAGCCGAC 3225  
Db 17267 AAACCAAGCTGGCACAGTGGGCAATTGCTGTGGGCTTGGGCAATGGCTTACCC 17208  
Qy 3226 GCGGAGCCCTTCCAAATTCGTTATTCATGTGGGCGACAGTATCCCGTGCMAAC 3285  
Db 17207 GGTCTACCTTTTCCAGATATCTATGTGCTCGCGGGGCGCTATACCGGTGMAAC 17148  
Qy 3286 TACGGTATGTCAAATTCGCGCGGCGCATGCGGCGTGAAGAGGCTTATATTATCAC 3345  
Db 17147 TACGTATCTGATCAGACAGATCGAAGAGGAGATTCGGAATTTTACTTTCAGC 17088  
Qy 3346 CCGTTCAGACACGTTTGTGTAAGT---CTCCGATCAAGGATGAGCGCGGACGAC 3402  
Db 17087 CCGGACAGAGTGTCTTATGCGGATGACACAGATGAGCTGAACAGTGTGTTTCA 17028  
Qy 3403 GTTGGCAAACTTGACGCTGTTGATGAGAGCGGCTTCACTCTGTTCTGGGCGAG 3462  
Db 17027 GCGGACCTTAATCAGTCTATTCGATGCGGCGGCAATTCACCTTTATATGCTGAT 16968  
Qy 3463 ATCGAGCGCATGAGTGGCTGTATG---ATGCTGTCGCGAATTTTGCCTGTGAG 3519  
Db 16967 ATGGGAGTATCTCGCATTTATGAGGACAGGAGACTTTCGATGATGAA 16908  
Qy 3520 GCGGATATATGCGCAGCTCTGATGAGAGAGCGGCTTCTGCAACATCGGCTGT 3579  
Db 16907 GCAGGACTCTGCAATTTTACTGGAAGAAATGCGCGCTTACGATTTGGGCGTGC 16848  
Qy 3580 CCGGTGGGCAATTTCAATTTTGAACAGTTCGCGGCTTCTGACCTGCGAATTCGAC 3639  
Db 16847 CAATTTGGGATGCAAGACTTTTCGCTGTGGCATTCGATTTCAATTTGCGCAATCAT 16788  
Qy 3640 GTTTAGTGCAGGATGCTGGCGGCGGAGTGAACCGCGGCAAGTTCAGGCTGTAG 3699  
Db 16787 CCGTATGCTCATTTGACCGGTGAGGCGGCGGATGAGGCAAGGCGCAAGTCTGACGA 16728  
Qy 3700 CTGCGTCAAGATTCTTCAACGAGGCGCGCACGAGCGGCGGCTTCCGCGCGCGAG 3759  
Db 16727 TTGCTGCGGATTTTCTCCACTATGAAACGGAAGAAACGCTGCGCGCTGACAG 16668  
Qy 3760 CAGCACTTTCGAGTATGCTTGGCACTTTGAGAGCAAACTACCCGAGTACATGAG 3819  
Db 16667 CAGAGCTAACAAAGATGCCATGTGCGCGGCTGCGTCAAGCACTGCTACTATATGAG 16608  
Qy 3820 CCTAAGCTTGTGAGAGCTGATGCTGCGGCTGACGCTTCAACGCAAGGCTGATCT 3879  
Db 16607 CCGAGTATCTATGTTAGCAGCCGATTTTCCGTTTACCGCTTAAAGGCAAGCTAGTCCG 16548  
Qy 3880 AAGGCGCTGCGGAGGAGATACCTGTGCGCGGCAAT-----TGGGGCGACAG 3933  
Db 16547 CAAAATTACAGCTGAGGCGCAAAATTTGCCACACAGCTGAGCGGCTGGGTCAATTC 16488

QY 3934 GCGCCACGGACGCTTGGAGAGATCCTCGTCGCGTCAAGAGAGTGTCCGGCGT 3993  
DB 16487 CAGGTGACAGTGGCTTACACAGCGGCTGTGGCTCTGCGACAGAGTACTCGGCGTG 16428  
QY 3994 GAGGTGCTGGGCTTCCAGACAGCTTCTGTCATCTTGTGTGACATTCATTCATCTT 4053  
DB 16427 AGCCACGTGTGCGCCGAGACGATTTCTCTGCTGGGGGCGCATTTATAGAAATTTGTTG 16368  
QY 4054 CGCATGAGAGGCTTTCGACAGAGGCTCGATAGGAGATGCGCATCAGCGAGTTGTC 4113  
DB 16367 CGTATTCAGCAGGACCTTGAGGCGATTTATCGGCGACAGAGATTCGATGTGATCTGTTC 16308  
QY 4114 CAGTACCGGAACTTCGCTCGCTGCG 4140  
DB 16307 CGTCCGCAACATCGCGGATGTGCG 16281

RESULT 7  
ABST7887  
ID ABST7887 standard; DNA; 32160 BP.  
XX AC ABST7887;  
XX DT 17-DEC-2002 (first entry)  
XX DE E. coli CFT073 genomic sequence #54.  
XX KM Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;  
XX KW urinary tract infection; open reading frame; ORF; uropathogenic;  
XX KM antibacterial; atropathic; nephrotropic; gene; ds.  
XX OS Escherichia coli.  
XX EN WO200259320-A2.  
XX PD 01-AUG-2002.  
XX PF 19-OCT-2001; 2001WO-US46833.  
XX PR 19-OCT-2000; 2000US-242412P.  
XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX PI Blattner FR, Welch RA, Burland VD;  
XX DR WPI: 2002-691532/74.  
XX PT New DNA sequences of the pathogenic Escherichia coli CFT073 strain,  
XX PT useful for preventing or treating E. coli CFT073 infection in humans or  
XX PT livestock  
XX PS Claim 1; Page 166-184; 765pp; English.  
XX CC The present invention relates to polynucleotide sequences from the  
XX CC genome of the pathogenic Escherichia coli strain CFT073. Almost all  
XX CC the sequences present in E. coli CFT073 are absent in the previously  
XX CC sequenced laboratory strain K-12. The polynucleotide sequences of  
XX CC the invention are useful for preventing, diagnosing or treating  
XX CC E. coli CFT073 infection in humans or livestock. The polynucleotide  
XX CC sequences are useful for preventing urinary tract infections and  
XX CC pyelonephritis. Likewise, the polypeptides encoded by the different  
XX CC open reading frames (ORF1-5) are useful for generating a vaccine  
XX CC against uropathogenic E. coli strains. ABST78834-ABST79085 represent  
XX CC genomic sequences from E. coli strain CFT073.  
XX SQ Sequence 32160 BP; 7001 A; 9686 C; 8766 G; 6688 T; 19 other;

Query Match 9.2%; Score 390.4; DB 24; Length 32160;  
Best Local Similarity 51.2%; Pred. No. 3e-73;  
Matches 1086; Conservative 0; Mismatches 997; Indels 38; Gaps 6;

QY 761 CGATGATTACTGGAAGGCGCATCGCGAGCTCCACCTCGCGAGCTTCCGATGA 820

DB 10021 GCGCGCTTACTGCGTGGCGAAGCATGACGCTTCCCCCGCGCGCTTCTTGGCGCTGG 10080  
QY 821 AGCGCATTCATACCTGTAAGAGATCCGCTTCCGCGACAGAGCAATGGCGCGT 880  
DB 10081 CTTGGAAACCGCCACGCTATATGTAAAGTCTCTTAATACCGACCGCGCATGATTTGCGG 10140  
QY 881 CGSACTCTGGGGTTCATTTGAAGCGGCGTGTGGGAGACGGCGCTGACCCCGAGCGGG 940  
DB 10141 CAACACGTGGACCGCTTTATGCAACCGGGCGGGAGATATGGCGTGAACCGCGAATGG 10200  
QY 941 TCATCTGGCTGCAATTTTCCAGAGTATCGGCGCTGGAGCGGACCCCGGTTTACGC 1000  
DB 10201 CGCTGGCCACCTGTTTTTCTCCGCTGGCTGCGTGGGGGGGCGCTGACGCGCTGCTGC 10260  
QY 1001 TCACATTAAGCTGTTCACCGGCTCCCGCCGTCATCCGCGGTGAAGATATACCGGGG 1060  
DB 10261 TTACATCATCTTATTCACCGCCGACCGCTGACCCGCGGTGGCGCGATGCTTGGCG 10320  
QY 1061 ACTTCACGTGATGGTCTCTCGACATCGACACCATCGCGACMAAGACTTCGAACAGC 1120  
DB 10321 ACTTCACCAATATCTTCTGCTGACACCGCTGATGGGATACCGTCAGCAACCTGG 10380  
QY 1121 GCGCTAAGCGTATTCAGAGCAGCTGTGGAGCGAGATCATCTGCGATTAACGGTA 1180  
DB 10381 CGCGTAA---AAACAGCTCACGTTTACGAGAGACTGGAGACATCGCCTGCTCGGCG 10437  
QY 1181 TCGAGTTCACGAGAGGCGCGCGGCTCTGCGGATTCMAACGAGCGCATTTGTTCCCG 1240  
DB 10438 TCGAATTAATCTCGTAACTCAAGCGCAGCGCTACCCCGCGCGC-----CGCGG 10491  
QY 1241 TGGTGTACGAGCGCGCTTAACGAGCAAGTGTGTGTCACTCTGTTCGAGAGCTCG 1300  
DB 10492 TGGATTTTACAGCAATTTGGGGGCTTCCCTTCACAGACCGCGCAATTCGCGGTTGG 10551  
QY 1301 GAATCCGGTGTACACGACGACGACACTCTCAAGCTGCTGTGATCATGACTCTACG 1360  
DB 10552 GCGAGCCGGAATGGGGCATCTCGCAACCGCGCAGAGTGTGATAGATCATCTGCGGTTGG 10611  
QY 1361 AGCAGATGGGGACTCTCTCGCTGGGAGCATCTGTGAAGAGAGTGTCCCGCCGAC 1420  
DB 10612 AGCATACAGGAGGAGTCTGCTACMAATGGACAGCAAGACGCGCTGTTCCTCGGGGCT 10671  
QY 1421 TTCTGACGACATGCTCGAAGCGTACGCTGTTTTTCTCCGCGGCTCATAGGAACAT 1480  
DB 10672 TAGTCGAAACATTTGTCAGCCCTACTGCGCAGTTGATTAACCACTCTGCGATGCGAAA 10731  
QY 1481 GGGGTGAACAGGTGC-----GCTGTTCGCTTCCGCTGCCAGCTAGAAGCGCGGCG 1532  
DB 10732 GCGCTGGCAAAAGCGCTTGGCAAGTATATATGCGCGCAGCGCGCGATAGCGCAAC 10791  
QY 1533 GAGCGCAAAACGCAACACGCGCTGTGAGGAGCATACGTCGACGCGCTTTCGCGGC 1592  
DB 10792 GGGATCAACKCCACCGGCTGCCCCCATTTCCGAAAGCTTGTGATGAAGCAATTTTCCG 10851  
QY 1593 GCGGCTGAGAGAGCTGCCCATGACGCTGCGCTGTGGCGGCGCAAGAGCTTCACGTA 1652  
DB 10852 TATGCTTGTGAACACCGCGCGCGCTGCGGTACGACATGCTTATCACTGGAATTA 10911  
QY 1653 CGAAGAGCTTTGCGCGCTTGGCGGCACTTGGCGCGCGGCTGCGCGACAGGGGCGACG 1712  
DB 10912 TCATAGAGTGAACAGCTATGCGCGCGCGCTTGGCGGCGCGGTTAATGAGTGGGGTTCA 10971  
QY 1713 CCGCAACCATTTGGTCTGCGGTGTGTATGAGAAAGAGCTGGAAGCAAGTGTTCGCGCTCT 1772  
DB 10972 GCCCGCATATATGTGTGTATCAAGATGTCCAAAGGCGCGACAGCAACATTTGTCGCTCT 11031  
QY 1773 GCGGCTGTGAGTCAAGCGCGCGCTTACGTCGCTGATCGGACCTACCGCGCGAGCG 1832  
DB 11032 GCGCTCTGCTGCGCGCGCGGCTTTACGTTTCGCTTTCGCTGATCAGCTTCGCGCACG 11091  
QY 1833 TATCCATCACTCTCTGATCATGTGTGAAGCTGCTGTGACGACGATGGCTGGA 1892

```

Db 11092 GGGCGAGAAATCTACGCTGACGCCAGCGCTCCGCTGCTCATTTGTTCAGACAGCC 11151
Qy 1893 TGGCAAACTGATGCGCGCCCGGGATCCAGCGGCTGCTGTGAGCGAGCCGGCTCGA 1952
Db 11152 CAGCGCGGGGTCA-----GACGATATTCGCCCTTGCTGGGACAGAGCCATT----- 11200
Qy 1953 AGGCGACGGCGACCGACCTCCGATGATGCGCCATTCAGACACTTGGGATCTGCGGTATG 2012
Db 11201 ---GAGGCGAGCGCGATGCGCAACCGGATGATGCGCGCCCGACGCAACCGGCTTACAT 11256
Qy 2013 CATCTACACTCGGGATCCAGCAAGGTGCGCCAAAGGGGTGATGATGATCATCGGGTGC 2072
Db 11257 TATCTACACTCGGGTCTACCGGATACGCGGAAGGGGTAGTATTTCTCACCGGGGAC 11316
Qy 2073 CGTCAACACCATCTGGAATCAACGAGCGCTTCGAATAAGGCGCCGAGACAGGTGCT 2132
Db 11317 GCTTAACACTGTGGATATCAATACCGCTATCAGTTGGCGCCCATGACAGGGTGTCT 11376
Qy 2133 GGGGCTCTCTGCTGATGCTGATCTCTGCTGATGATGATGATGATGATGATGATGATG 2192
Db 11377 GGGGCTCTCTGCTGATGCTGATCTCTGCTGATGATGATGATGATGATGATGATGATG 11436
Qy 2193 GGGGCGGTACGATCGTGTGCGCGGACCGGTCAGCTGCAAGCTGCGCATCCGCGCATTTGGG 2252
Db 11437 GGGGCGCGCGCTGTGTATGTGTATGTAAGAAATCAACGCGCGCATCTCAAGCATGTGTGA 11496
Qy 2253 GTTATCGAAGAGAGAGAGTGAAGTGTGTGAATCTCGGTGCGGCGCTGATGCGATGCT 2312
Db 11497 GCTGATCCAGCCCATCAGGTCAAGTCAAGTCAAGCAAGCTCGGCGCTTTCGATGATGCT 11556
Qy 2313 CGTCAAGCATTTTGGAGGCTGCGCGCGGATTCGCTGCTGATGCTGCTGCTGCTGCTGCT 2372
Db 11557 GCTGATCGTGTGTGAAGTTTGGCGGACGCGACGCGGAAACCTGCGCGCATGTGCTGCT 11616
Qy 2373 GAGCGCGCATGATGATCCCGGTGCGCTGCTGCGCATGCTCAAGCCATCAGCGCGCGCT 2432
Db 11617 TTCGCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11676
Qy 2433 GTGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2492
Db 11677 ACAATTTATCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11736
Qy 2493 GAGGAACCTGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2552
Db 11737 TCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11796
Qy 2553 GTTCCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2612
Db 11797 CTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11856
Qy 2613 CATTTGGGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2672
Db 11857 GATTTGGGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11916
Qy 2673 CTTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2732
Db 11917 ATTTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11970
Qy 2733 CTTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2792
Db 11971 CTTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12030
Qy 2793 ATTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2852
Db 12031 ATTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12090
Qy 2853 CGGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2873
Db 12091 AGTACCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 12111

```

RESULT 8  
ABX34289

```

ID ABX34289 standard; DNA; 135638 BP.
XX AC ABX34289;
XX 11-FEB-2003 (first entry)
DT S. atroliivaceus leinamycin biosynthesis gene cluster.
XX DE
XX S. atroliivaceus leinamycin biosynthesis gene cluster.
XX leinamycin biosynthesis gene cluster; lnm: open reading frame; ORF;
XX anti-tumour antibiotic; broad spectrum antimicrobial activity;
XX Gram-positive; Gram-negative bacteria; chemical modification;
XX metabolite; apo-carrier protein; holo-carrier protein; tumour;
XX polypeptide; hybrid polypeptide/polypeptide metabolite; lnm production;
XX cytosolic; gene; ds.
XX OS
XX Streptomyces atroliivaceus.
XX PN
XX MO200271179-A2.
XX PD
XX 03-OCT-2002.
XX PF
XX 22-MAR-2002; 2002WO-US08937.
XX PR
XX 26-MAR-2001; 2001US-278935P.
XX PA
XX (REGC ) UNIT CALIFORNIA.
XX PA (KYO) KIOWA HAKKO KOGYO KK.
XX PI
XX Shen B, Cheng Y, Tang G;
XX MPI: 2003-018907/01.
XX P-PDB: ABU11341, ABU11342, ABU11343, ABU11344, ABU11345, ABU11346,
DR ABU11347, ABU11348, ABU11349, ABU11350, ABU11351, ABU11352, ABU11353,
DR ABU11354, ABU11355, ABU11356, ABU11357, ABU11358, ABU11359, ABU11360,
DR ABU11361, ABU11362, ABU11363, ABU11364, ABU11365, ABU11366, ABU11367,
DR ABU11368, ABU11369, ABU11370, ABU11371, ABU11372, ABU11373, ABU11374,
DR ABU11375, ABU11376, ABU11377, ABU11378, ABU11379, ABU11380, ABU11381,
DR ABU11382, ABU11383, ABU11384, ABU11385, ABU11386, ABU11387, ABU11388,
DR ABU11389, ABU11390, ABU11391, ABU11392, ABU11393, ABU11394, ABU11395,
DR ABU11396, ABU11397, ABU11398, ABU11399, ABU11400, ABU11401, ABU11402,
DR ABU11403, ABU11404, ABU11405, ABU11406, ABU11407, ABU11408, ABU11409,
DR ABU11410, ABU11411.
XX
XX Novel gene cluster responsible for synthesis of leinamycin in
PT Streptomyces atroliivaceus useful for making various peptide and/or
PT polypeptide, and/or hybrid polypeptide/polypeptide metabolites
PS
PS Claim 6; Page 81-127; 185pp; English.
XX
XX The present invention relates to the isolation of the Streptomyces
XX atroliivaceus leinamycin (lnm) biosynthesis gene cluster containing
XX 71 open reading frames (ORFs) (ORFs -35 through -1, ORFs lnmA through
XX lnmI, and ORFs +1 through +9). Leinamycin is a novel anti-tumour
XX antibiotic produced by several Streptomyces species. It exhibits
XX broad spectrum antimicrobial activity against Gram-positive and
XX Gram-negative bacteria, but not against fungi. The polypeptides encoded
XX by the lnm biosynthesis gene cluster ORFs are useful for chemically
XX modifying a molecule in a host cell. The host cell is a bacterium or
XX eukaryotic cell, including a mammalian, yeast, plant, fungal, or insect
XX cell. The molecule is an endogenous metabolite produced by the host
XX cell or exogenously supplied metabolite, or an amino acid, and the
XX polypeptide is a peptide synthetase or amino transferase. The
XX polypeptides encoded by the lnm gene cluster are useful for converting
XX an apo-carrier protein to a holo-carrier protein. lnm shows potent
XX antitumour activity in tumour models in vivo. The lnm gene cluster
XX modules and/or catalytic domains are useful for making various peptide
XX and/or polypeptide, and/or hybrid polypeptide/polypeptide metabolites.
XX The proteins encoded by the ORFs are useful alone, or in combination
XX with other active domains to modify various target substrates. The
XX lnm gene cluster is useful to upregulate endogenous lnm production to
XX permit lnm production in cells and/or to make various modified lnm.
XX lnm, its analogue, or other polypeptide, peptide or hybrid
XX polypeptide/peptide metabolites are useful as therapeutic agents, to

```





CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (A199693) and  
CC H37Rv (A199698). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at [seqdata.uspto.gov/sequence.html?docID=6294328B1](http://seqdata.uspto.gov/sequence.html?docID=6294328B1).  
XX  
XX  
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
  
Query Match 8.5%; Score 358.2; DB 22; Length 4411529;  
Best Local Similarity 48.4%; Pred. No. 6.6e-66;  
Matches 1252; Conservative 0; Mismatches 1283; Indels 54; Gaps 7;  
  
QY 286 CACGCTATCGCAATACGACCTGTACGATCTGACGTCGCGGAGGCTGAGCGGCTTT 345  
DB 2675463 CACCTTACGTGCGAGTTCCATGTGTCGCCGGTGTGATCCGACAGGCTCCGCGGCGGCC  
QY 346 CGGAAAGTCGTGCGCGGACGACGATGCTTCGCGGCCACACGCTGCGGACATGATGACG 405  
DB 2675403 ACCCGGTGGCGCTGCGGACCGGATGCTGCGGATGCAATTTGCTCCGACCGGACCGG 2675344  
QY 406 GTGATGACCTTAAAGTTCAGCGCCGACATGAGATGATGATGTCGCGGCGCTCGACCGG 465  
DB 2675343 CGCATCCCGCGCGCGGAGATTCGCGACTTCCCATCAATGATTTCCGACCTGCTGAC 2675284  
QY 466 AGCAGCGGAAAGCGAGGCTGCTGTGTCGATGCGATGTCGACCGACCGCATCTATGAC 525  
DB 2675283 GTGCGCCCGGATGCTGATGACGCGGTTGCGGAGATCCGACGCGCAATGACCGG 2675224  
QY 526 ACCGAGCGGCTTCGCTCTATCACTGCTGCGCGCTTCGCTGACGCGGACAAACCGGT 585  
DB 2675223 CAGCTGACGCTGCGGATTCGAACTTGGCTGACGTTATACCGGAGGACCGACCGCG 2675164  
QY 586 CTGCTGCTAGATGATGATCTCATTAAGTTGACCTTGGACGCTGCTCATCTTTCAAG 645  
DB 2675163 CTACATGTGACCTGACATGACGCGCGTGAAGCGCATGATGACGCACTTTGCTGCGC 2675104  
QY 646 GACTGCTGAGCTTCTACGAAAGATCCGAGACCTCTCCCTGCTGAGGCTCTGCTAC 705  
DB 2675103 GACTGCGGCGCTCTATGACGCGCTGACCGCGGCTGACCTTACCGGCGG 2675044  
QY 706 CGGATTAATGTACTCGCGCTGAGTCTCGCAAGATCTGAGCGGATCAACGATGATG 765  
DB 2675043 TACCGGACGCTATCGAGGCGGAGAGACGCTGCCCAACCGGTTGCGGACCGGACG 2674984  
QY 766 GATTAATGGAAGGCGGATCGCGGAGCTCCACCTCCGCGGACGTTCCGATGAAAGGC 825  
DB 2674983 GACTGTGAGCGGAGGCAATCCGCGAGCTGCGGATTCGCGCTGCTTACCGGCGC 2674924  
QY 826 GATTCATTAACCTTGAAGAGATCCGCTTCGCGGACACGAGGATGAGTCCGCTGAGAC 885  
DB 2674923 GCGGCGGAACGCAACCGCGCGGACGACCGCGGCG---TGGCACTGGCTAAGACCGGAG 2674867  
QY 886 TCTGGGCTGATTAAGAGCGGCGTGTGCGGAGCGCGGCTGACCCCGACGCGGCTATC 945  
DB 2674866 ACCCGGACGCGCTGTGTCGCGGAGCGCGGCGCGGCTGACCCCGGAGATGACGCTG 2674807  
QY 946 CTGGCTGCAATTTCCGAGATGATCGGCGCTGAGAGCGGACCGCGGTTTACGCTCAAC 1005  
DB 2674806 GCGCGCGCTTTGCGCAACGCTGCGCGCTGTGTCAGCGCTGTGCGGTTCTGCGAAG 2674747  
QY 1006 ATTAACCTCTTCAACGCGGCTCCCGCTCATCCGCGGCTGAACGATTAACCGGAGCTTC 1065  
DB 2674746 CTGCGCTTTGCTAGTGGCCAGGCGCTGCTCATGACAGTGCACCTGCTGCTGAGCTTC 2674687  
QY 1066 AGCTGATGCTCTCTGACATGACACCACTCGGACAAAGGCTTGAACAGCGCGCT 1125  
DB 2674686 ACCTCTCGCTGTGCTGACGCTGATGACCGGCTGCGGACCGGCGGCGGCGGCG 2674627

QY 1126 AAGCTATTCAAGACACTGTGCGGAGACGATGATCTGCGAGCTAAGCGGTATCGAG 1185  
DB 2674626 CAGGCGGTGCGAAGAGCCCTGTGCGAGCCCGGACGACAGTACATCCCGGCTGTCT 2674567  
QY 1186 GTCCAGGAGAGGCGCGGCTCTGCGGATTCACAGGCGGCAATGTTCCCGCTGCTG 1245  
DB 2674566 GTGCTGCTGACCTACGCGCGGACCGGTGCGACCCAGGTGCTGACCGGTATTCAC 2674507  
QY 1246 CTCACGAGCGGCTTAAACGACGATGCTGTGCTACCTGTTGACAGAGCTCGGAAT 1305  
DB 2674506 AGCGCGCTGGGCTCGGCGACCTTTTCTGCCGAGCTGACCGAGCA---ATTCCGACA 2674450  
QY 1306 CCGGTGATACCGAGACCGGACCTCTGAGCTGCTGATCATGATGATGATGATGATGATG 1365  
DB 2674449 CCGGATGATATTTTCCGAGGCGCCGAGGTGCTGACCGCCAGGTACCGAGTTTC 2674390  
QY 1366 GATGGGAGACTTCGCTCTGCGGTCGAGACATGCTGACGAGAGTGTTCGCGCGGACCTTCTG 1425  
DB 2674389 GACGCGGTGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 2674330  
QY 1426 GACGATGCTTGAAGCGTACGCTGTTTCTCCGCGCT-----CACTGAGAACCA 1479  
DB 2674329 GACGCTATGTTACCCACAGGTGACGCAATGCTCGGTTGGCCCGCGGAGACAGCG 2674270  
QY 1480 TGGGGTGAACAGGTGCGCTTTCGCTTCGCGCTGCGGCTGCGGCTGAGAGCGGCGGCGCA 1539  
DB 2674269 TGGGATGCGCGAGCGCTGCGGCTGACCGCGGCAACCGCGGTGCGCGCGCTG 2674210  
QY 1540 AACGCGAACAGCGCTGCTGACGAGATACGCTGACGCGCTGCTGCGCGCGGCTG 1599  
DB 2674209 AAGGATGCGACCGCGCGCGCGCGCGGACGAGGCGTGCAGACGCGGTTTTTCCGACGCG 2674150  
QY 1600 GAGCAGTCCCGCATGACAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1659  
DB 2674149 CAACAGAGCGCGGACGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2674090  
QY 1660 CTTTGGCGCGTTCGCGGCACTTGGCGCGGCGGCTGCGGAGCAGGCGGCGGCGGCAAC 1719  
DB 2674089 CTGCGGACAGGACATCGCGGCTGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 2674030  
QY 1720 ACATTGCTGCGGCTGATGAGAGAAAGCTGCGGAGAGGCTTTCGCGGCTGCTGCTG 1779  
DB 2674029 GACACGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673970  
QY 1780 CTGAGTCAAGCGCGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1839  
DB 2673969 TTGGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673910  
QY 1840 TACCTCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1899  
DB 2673909 CCGATCTGCGGACCGGCTTCCGATCTTAAGCGCT-----CGTTGCGG 2673865  
QY 1900 CTGTATGCGCGCGCGGATTCAGCGGCTGCTGCTGAGCGAGCGCGGCTGCAAGGCGAC 1959  
DB 2673864 CCGCATGCAAGTACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673805  
QY 1960 GCGGACAGCGCTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2019  
DB 2673804 GCGCGGCGAATTCGCTGCGCGGCTGAGCGATTCACCGCGCTGCTGCTGCTGCTG 2673745  
QY 2020 ACCCTGGGATTCACAGGCTGCGGAGGAGGATGATGATGATGATGATGATGATGATGATG 2079  
DB 2673744 ACCCTGGGCTCAACCGGGAACCAAGGCTGCGAGGTGCGGACGACCGCGGATGAGAC 2673685  
QY 2080 ACCATCTGACATCAAGAGGCTTTCGAATAGGCGCGGAGACAGAGGCTGCGGCTC 2139  
DB 2673684 ACCGAGGACCTTATTCGAGCTTTCGAGCTGAGCGCGGAGATGCTGCTGCTGCTG 2673625  
QY 2140 TCTGCTGAGCTTGTGATCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 2199  
DB 2673624 GCGAGCTGAGATGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2673565  
QY 2200 AGATCTGCTGCTGCGGAGCGCTTCAAGCTGCGGATTCGCGGCTTGGCAGAGTTGATC 2259

```
Db 2673564 GCGATCGGTGTGTGCGAAGACGCGAGCGCGCGATCCCGACGCTCGGCGCCGCGCTTATC 2673505
Qy 2260 GAACGAGAGAGGTGCGGTGTGGAACCTCGGTCGGGGCGGTGATGCGATGCTGCTGAG 2319
Db 2673504 GACACTTACGAGGTACGCGCTGTGATTTTCAATCCGGGCTGTGTGACATGTGCTGAA 2673445
Qy 2320 CATTTTGAAGGTTCGCGCGATTCGCTCGCTAGTCTGTGCGCTTTCGCTGAGCGGC 2379
Db 2673444 GTCCGCGGGGGCGCGCTGTGCTGCG-----CTGCGAGCGGTGCGCGTGGCGGC 2673397
Qy 2380 GACTTGATCCCGGTGGCTGCTGCTGCGAAGCTCCAGGCCATCAGGCCCGGCGTGTG 2439
Db 2673396 GACTGGGGTCCTCCGACCTGCGCCGCGCTGCGAGTCAAGCCCGACGCGATTC 2673337
Qy 2440 ATCAGCCTGGGGGGGCGACGAGCGTGTGATCTGTGCTCATGCGGTACCCGCTGAGAAC 2499
Db 2673336 GCGGGGTTGGGTGAGCCACCGAAGCCGGGTCAAGCAACATTTTGAAGTCCAGAAC 2673277
Qy 2500 GTGACCTTATCG-----TGGCGAGCATCCCTTACGCGCGCTCGCTGCGAACGAG 2550
Db 2673276 GCGGCGATCTGCGACCGGACTGGGCTCTCGGTCCATACGGGTCCCTTCCACAAC 2673217
Qy 2551 AGCTTCCAGTCTGATGAGCGCTCGAACCGCGCGCTGTGCTTCCGGGCGACTC 2610
Db 2673216 GCTGCGGGGTAGTGCGCGACGCGCGACGACTCCCGATGGGTGCGCGTGAATTG 2673157
Qy 2611 TACATTTGGCGGGGTGGGCTGCGACTGGGCTATCTGCGCGCATGAGAGAAACCGCGAG 2670
Db 2673156 TGGGTGCTCGGTGCGGAATGCGCCGGGGTTACCTGTGCGCTCCGACCTGACCGCGAG 2673097
Qy 2671 AGCTTCTCTGTGCAACCCGAGACCGGGAGCGCTCTCAAGACCGGCGATCTGGGCGC 2730
Db 2673096 CGCTTCGTGAGC-----ATGACGCGCGCACCTGTGTATCGACCGGTGATCTGCGCGC 2673043
Qy 2731 TACCTGCGCATGGAACATGAGTTTCAATGGGGCTGAGAGCAACCAATCAAGCTTGC 2790
Db 2673042 TACTGCGACGACGCGACCTGGAATTCGTGCGCGCTGCGCATCACCGGTCAAGTACG 2672983
Qy 2791 GGATTAACCGGCTGAGCTGGGGAAATCGAGAGAAAGCTCAAGTCCATCCGAAGTACG 2850
Db 2672982 GGGTACCGCGTGAACCTGCGGAGATCGAAGCGCGCTGACGCGCTTCCCGGTGTGAC 2672923
Qy 2851 GACGCGGTG 2859
Db 2672922 GCGGCGGCG 2672914

RESULT 10
ID AAI99683 standard; DNA; 4403765 BP.
AC AAI99683;
XX
XX 15-JAN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX variation; epidemiology; patient treatment; epidemic monitoring; de.
XX
XX Mycobacterium tuberculosis.
XX
XX US6294328-B1.
XX
XX 25-SEP-2001.
XX
XX 24-JUN-1998; 98US-0103840.
XX
XX 24-JUN-1998; 98US-0103840.
XX
XX (GENO-) INST GENOMIC RES.
XX
```

```
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
DR
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX determining the nucleotide sequence of the strain at positions in the
XX genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ
XX
XX Claim 4; SEQ ID NO 2; 3bp + Sequence Listing; English.
XX
XX The invention relates to evaluating strain variation within and between
XX different populations of the tuberculosis bacterial pathogen,
XX Mycobacterium tuberculosis or related Mycobacterium by determining the
XX nucleotide sequence of the first strain at positions in the complete
XX sequence of the genome that correspond to positions that differ in the
XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
XX H37Rv (AAI9682). The method is useful for evaluating strain variation of
XX M. tuberculosis and has valuable application in the fields of
XX tuberculosis genetics, epidemiology, patient treatment and epidemic
XX monitoring.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
XX Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
```

```
Query Match 8.2%; Score 349.2; DB 22; Length 4403765;
Best Local Similarity 50.9%; Pred. No. 5.5e-64;
Matches 1073; Conservative 0; Mismatches 983; Indels 54; Gaps 8;

Qy 765 GGATTAATGGAAGGCGGCGATCGCGGAGCTCCACCTCCGCGAGCTTCGATGAAGC 824
Db 2672280 GGAATGTTGGGCGGACGCGATCCGCGAGTCCCGGCTCCGCTTACCCGGGC 2672221

Qy 825 CGATTCATTAACCTGAAGAGATCCGCTTCGCGACAGGAGCAATGCTCCGTCGA 884
Db 2672220 CGGCGGAGACGCGACCGCGCGCGCGACGACCGCGCGC---TGGCACTGTGATGACCCGCA 2672164

Qy 885 CTCTGGGTGATGATGAAGCGCGCTTCGGGAGCGCGGCGTGAACCCCGAGGGGCTAT 944
Db 2672163 GACCGGCGACGCGTGTGTCGCGCCAGCGCGCGCGCGCATACCCCGGAGTACGCT 2672104

Qy 945 CCTGCTGATTTTCCAGGTGATCGGCGCTGAGCGCGAGCCCGGCTTACGCTCA 1004
Db 2672103 GCGCGCGGCTTTGCGCAAGTGTGCGCGCTTGTGATGCTGCGCTTCTGCTAA 2672044

Qy 1005 CATTAACGCTTTCAACCGGCTCCCGCTCATCGCGCGTGAACGATATCACCGGGACTT 1064
Db 2672043 CCGCGGCTTTGATGATGCGCGCGCGCTCATGACGATGACGCTGCTGCTGATCTT 2671984

Qy 1065 CACGTGATGCTCTCTGAGCATGACACCATCTCGGAGCAAGAGCTTCAAGCGCGC 1124
Db 2671983 CACCTCTCGCTGTGCTGCGACGTGATCTGACCGGTGCGGACCGCGCGCGCGGC 2671924

Qy 1125 TAAGGTATTAAGAGCAGCTGTGGAAGCGATGATCACTGCGAGCTTAACGATGCA 1184
Db 2671923 GCAAGCGGTGAGAGAGCGCTG-CGACGCGCGAGCGCACAGTGCATACCCCGGCTTC 2671865

Qy 1185 GGTCCAGAGAGCGCGCGCGGTCTCGGGATCAACGAGCGCGATTTTCCCGTGT 1244
Db 2671864 TGTGCTGTGATCACTCAAGCGCGCACCGGTGACACCGGTGTGACACGGGTGATTCAC 2671805

Qy 1245 GCTACAGAGCGCGCTTAACGAGCAAGTGTGTGTGATCACTCTGTTCGAGAGCTCGAAC 1304
Db 2671804 CAGCGCGGTGGGCTCGCGACCTTTTC--TGCCGAGCTTCAACGAGCAATTCGGAC 2671747

Qy 1305 TCCGATGATACAGAGCAGAGCTCTCAAGCTCTGATGATCACTCAAGCA 1364
Db 2671746 ACCCGATGATCACTTTCGAGGGGCCCGAGTGTCTGACCGCCAGTCAACGAGTT 2671687

Qy 1365 CGATGGGAGCTGCTCTCGCGTGGGACATGTCGACGAGATGTTCCCGCGGACCTTCT 1424
```

Db 2671686 CGACCGCGGTGCTGCTGTGAAGTCCGCGAGGGGGTCTTTGACACCGCGCTCAT 2671627  
Qy 1425 GGAACGATGCTGGAAGCGTACGTGTTTCTCGGCGGCTCACT-----GAGAACCC 1478  
Db 2671626 CCACGCGATGTTTCAACCCACAGAGTGCAGAAATGCTCCGGTTGGCGCCGGGAGACAGCC 2671567  
Qy 1479 ATGGGGTGAACAGTGTGCGTGTTCGCTTCGCGTCCGCTGACAGTAAAGCGGGCGAGGCC 1538  
Db 2671566 CTGGAGTGGCGCGAGCCCGCTCGCGCTACCGCGCGCAACGCGCGGTGGCGCGCT 2671507  
Qy 1539 AAACGCGAACACGCGCTGCTGAGCAGACATACGCTGACAGCGCTGTTGCGCGCGCGGT 1598  
Db 2671506 GAACGCTGCGACCGCGCCCGCCACGACGAGAGGCGTGCAGACAGGTTTTCGCGCAGCG 2671447  
Qy 1599 CGAGCAGTCCCGCAGCTGCGCGCTGCTGCGCGCGCAGACGCTCACTGACAGAA 1658  
Db 2671446 CCAACGACGACCCGACGCGCCCGCGGTGTCGCAATTCGCGGCACTTACGACCCCA 2671387  
Qy 1659 GCTTTCGCGCGCTGCGCGGCACTTGGCGCGCGCGCTGCGCGAGCAGGGGGCAGCGCCGAA 1718  
Db 2671386 ACTGCGCGACACAGGCAATCGAGCGTGGCGCGCGCTGCTGCGCGGCGCTACGAGTCCG 2671327  
Qy 1719 CACATTTGCTGCGGTGTGATGAGAAAGCTGGAGCAGATTGTCGCGTTCCTCGCGT 1778  
Db 2671326 CGACACCGTCCGCGGTCTGGTCCGAAACGCGCGAACAGTGGCGGTGTGCTGGGANT 2671267  
Qy 1779 GCTCGAGTCAAGGCGCGGCTTACGCTGCGATCGATGCGACCTACCGCGGAGCGATCCA 1838  
Db 2671266 TTTGGCGCGCGCGGTCTATCTCGCATCGCGCTGACAGCGCCCGCGACCGCGCGA 2671207  
Qy 1839 CTACCTCTCGATCATGTGTGTAAGCTCGTGTGCTGACGACCGCATGCGTGGATGGCAA 1898  
Db 2671206 GCGCATCTCGGCGACCGGTTCCGTCACTT-----AGCGTCTGTTGCGG 2671162  
Qy 1899 ACTGTCATGCGCGCGCGGATCCAGCGCTGCTGTGACGAGCGCGCGTCCAAAGCGCA 1958  
Db 2671161 CCGGCGATCCAGATCGAGTGGGTGCGGTCCGACGCTGTTGCTGTGAGTGTGTCGCG 2671102  
Qy 1959 CCGGCGACGACCTCCGATGATGCGCATTCAGACACTTGGATCTGCGTATGCTACTTA 2018  
Db 2671101 CCGCGCGGAGAAATGCTGCCGCGGCTAGCGATCCACCGCGCTGCTATGTGTTGT 2671042  
Qy 2019 CACCTGGGATCCAGAGGGTGGCGCAAGGGGGTGAATGATCATCGGGGGTCCGTCAA 2078  
Db 2671041 CACCTGGGCTCAACCGGGAGAACCAAGGGTGTGAGTGGCGCAAGACCGCGATGAA 2670982  
Qy 2079 CACCATCTGGAATCATACAGAGCGCTTCGAAATAGGAGCCCGAGACAGGGTGTGCGCT 2138  
Db 2670981 CACCGTGAAGACCTTTCATCCGGGCACTTGAAGTGAAGCGCGAGATCGCTGCTGCCCT 2670922  
Qy 2139 CTCTCGCTGAGCTTGAATCTCTCGGTCTATGATGTGTTTGGGATCTTGGCGCGGCGG 2198  
Db 2670921 GCGGACGCTGAGTGTGAGATGCTGCTGAGCATCTTTCGCGCGCTCGCTCCGCGCG 2670862  
Qy 2199 TAGCATGCTGTTGCGGAGCGCGTCCAAAGCTGCGCGATTCGCGGCAATTTGGGAGAGTTGAT 2258  
Db 2670861 AGCGATGCTGTTGTTGAGAGAACGCAAGCGCGCATTCGCGACGCTGGGCGGCTTAT 2670802  
Qy 2259 CGAACGAGAAAGTGAAGTGTGAACTCGGTGCGCGCGCTGATCGAGTGTCTCTGCA 2318  
Db 2670801 CGACACTTACAGAGTCAACCGCGTGAATTTCAATGCCGGGTGGTTGACATGTGCTGCA 2670742  
Qy 2319 GCATTTTGAAGGTGCGCCGATTCGCTCGTGAAGTCTTGGCGCTTTCGCTGTGAGCGG 2378  
Db 2670741 AGTGGCGGGGGCGCGCTG-----TCGTGCTGCGAGCGGTGCGCTGCGCGG 2670694  
Qy 2379 CGACATGATCCCGGTGAGGCTGCTGCTGCGAGCTCAAGGCGCATAGGCGCGCGGTCCGAT 2438  
Db 2670693 CGACTGGGTGCTGCTCCGACCTGCGCGCTGCAAGTGTCAAAGCCCGAGCCACGCTT 2670634  
Qy 2439 GATCAGCTGGCGGGGCGACCGAAGCGTGCATGTGCTCATCGGTACCCCGTGAAGAA 2498  
Db 2670633 CGCGGGGTGTGGTGGAGCCACCGAAACCGGCTGCACAGCACTTTTGAAGTCCAGGA 2670574

Qy 2499 CGTGCAC-----CTATCGTGGCGAGCATCCCTTACGGGCGCTCCGCTGCGAACCA 2549  
Db 2670573 CCGGCGCAATCTGCCACCGGACTGGGCTCGGTGCTCATACGCGCTCCGTTTCCAAACA 2670514  
Qy 2550 GACGTTTCAACGCTGCTGATGAGCGCGCTCGAAACCGCGCGTCTGGGTTCCGGGCAACT 2609  
Db 2670513 CGCTGCGCGGTGAGTGGCGGACAGCGGCGACGACACTGCCCCGATTGGGTGGCGGTGAGTT 2670454  
Qy 2610 CTACATTTGGCGGGGTGCGGCTGCGCACTGCGGCTACTGCGCGCATGAAGAGAACCGGCA 2669  
Db 2670453 GTGGGTGTCGCGTCCGGAATCGCCGGGGTTACCGTGGCGCTCCCGAGCTGACCGGA 2670394  
Qy 2670 GAGCTTCTCTGTCGACCCCGAGACCGGGGAGCGCTTCAACAAGCGCGGATCTGGGCG 2729  
Db 2670393 GCGCTTCTGTCGAGCATGACGCGCGCA-----CTGTATGCGACCGGTGATGCGCG 2670340  
Qy 2730 CTACCTGCGCGATGAAACATCATGATGAGGCGGTGAGACAAACCAATCAAGCTTCG 2789  
Db 2670339 CTACTGCGACGACGCGCACCTGAGATTGTCGCGCGTCCGATCAACCGGTCAAGATCAG 2670280  
Qy 2790 CGGATACCGCGTTGAGCTGCGGGAATGAGAGAAAGCTCAAGTGCATCCGACGTACG 2849  
Db 2670279 CCGGTACCGCGCTCGAATCGCGAGATCGAAGCCGCTGCGAGCGCTTGGCGGTGTGCA 2670220  
Qy 2850 CGACGCGGTG 2859  
Db 2670219 CCGGCGCGCG 2670210

RESULT 11  
AAS53964  
ID AAS53964 standard; DNA; 1416 BP.  
XX  
AC AAS53964;  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Klebsiella pneumoniae DNA for cellular proliferation protein #93.  
XX  
XX  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
XX  
XX antibiotic; antibacterial; drug design.  
OS Klebsiella pneumoniae.  
XX  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
XX  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT,  
PI Yamamoto RT, Xu HH;  
XX  
XX  
DR WPI; 2001-611495/70.  
XX  
XX P-PSDB; AAU36105.  
XX  
XX  
PT New polynucleotides for the identification and development of  
XX  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX  
PS Claim 27; Seq ID No 7601; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1416 BP; 271 A; 425 C; 429 G; 291 T; 0 other;

Query Match 7.2%; Score 305.6; DB 23; Length 1416;

Best Local Similarity 54.4%; Pred. No. 1.7e-55; Mismatches 559; Indels 21; Gaps 3;

Matches 692; Conservative 0; Mismatches 559; Indels 21; Gaps 3;

QY 1602 GCAGCTGCCCATGACGCTGCGGTGTGCGCGCGCAAGAGCTCAGCTACGAGAGACT 1661  
 DB 57 GCAACAGCGCGAGCGGCTGGCGGTAAACGACATGGGTTATCAGTAGAATTATCAGAGCT 116  
 QY 1662 TTCCGCGCGGTTCGCGGCACTTGGCGCGCGCTGCGCGAGCAGGCGGCAACG 1721  
 DB 117 GACACACTATGCGCCCGCTTGGCGCGGCGAGTTAATCGAGTCGCGGGTTCAAGCCGCGCA 176  
 QY 1722 ATTGTCGCGGCGGTGATGAGAGAGGCTGGAGAGAGGTTCTCGCGTCTCGCGTCT 1781  
 DB 177 TAATGTGGTATCAGATGTGCAAGAGCGCAGAGACAATCTGTGCGTTCGCGTCT 236  
 QY 1782 CGAGTCAGGCGCGGCTTACGTCGCGATCGATCCGACCTACCGCGAGAGCTATCAGTA 1841  
 DB 237 GCTGCGCGGCGGCGTTTACGTTCCGTTCCGCTGATCAGCCTCCGCGAGCGGAGAA 296  
 QY 1842 CCTCTCATCATGCTGAGAGTAAGCTGCTGCTGACGACGACCATGAGTGGCAACT 1901  
 DB 297 AATCTACCTGACGCGCAGGCTCGGCTGGTCTCA-----TTTGTACAGACAGACGCGAG 350  
 QY 1902 GTCATGCGCGCGGAGATCCAGCGGCTGCTCTGACAGCAGCGCGGCTCGAAGCGAGG 1961  
 DB 351 CGCGGCGTCAAGCATATTCCTGCTGCTGCGAGCAGCGGCAATGAGCGCGAGCGCAT 410  
 QY 1962 CGACGAGCTCCGATGATGATCCCATTCAGACACTTCGATTCGCGGTATGCTATCAGC 2021  
 DB 411 CGCCCAACC-----GATGATACGCGCGCCCGCCAGCAGCAGCGCTATATTTACAC 461  
 QY 2022 CTCGGGATCCAGAGGTTGCCAAGGCGGTGATGATCATCATCGGGGTGCGCTAACAC 2081  
 DB 462 CTCGGGCTCTACCGGTACGCGCAAGGGGTACTCATTTCTCACCGGGAGCGCTTAACAC 521  
 QY 2082 CATCTGACATCAACAGCGCGTTGGAATAGCGCCGAGACAGAGTCTGCGCTCTC 2141  
 DB 522 CTGTTGCAATCAATACCCGCTATCAGGTGCGCGCAGTGAAGAGGTGCGCTCTC 581  
 QY 2142 CTCGCTGAGCTTCATCTCTCGGTATGATGTTCGGGATCTCGCGCGCGGCGGTAC 2201  
 DB 582 CGCCCTACATTTTATTTATGCTTTACGACATTTTGGCGTACTCGCGCGCGGCGGAC 641  
 QY 2202 GATCTGTGTCGCGAGCGCTCAAGCTGCGCATCCGCGCATTTGGCGCAAGTTGATCGA 2261  
 DB 642 GCTGTGTATGTGATGAGAAATCAACGCGCGCATCTCTACGATGATGAGTATCA 701  
 QY 2262 ACGAGAGAGTGAAGGTGTGGAAGTCTGTCGCGCGCTGATGCGGATGCTCTGTGAGCA 2321  
 DB 702 GCGCATCAGGTACGCTCTGGAACAGCGTCCGCGCGCTGTCATGATGCTGTAACCTG 761

QY 2322 TTTGAGGGTGCAGCCCGATTCGCTGCTAGGCTCTTCGCGGCTTTCGCTGAGCGGCA 2381  
 DB 762 GTGTGAAGTTTTCGCGCGACGCGCAGGCAAGCAACCTGCGCGCATGATCTTTCGCGCA 821  
 QY 2382 CTGATCTCCGCTGCGGCTCTGCTGCGAGCTCAAGGCTATGAGGCCCGGCGTGTGAT 2441  
 DB 822 CTGATCTCGGCTTGAACCTCCCGCGCGCTTATCGGCTTTCGCGCGCAAGCAATTTAT 881  
 QY 2442 CAGCTGGCGGCGGCGCAACGAGCGTGAATGCTGATGCTGAGTACCCCGTGAAGAGCT 2501  
 DB 882 CGCGATGAGCGCGCGCACCGAGCGGTCTATCTGTATACCTCTCGCAATTCACGAGCT 941  
 QY 2502 CGACCTATCTGTGCGAGCATCCCTTACGCGCGTTCGCTGCGCAACGAGCTTCCAGCT 2561  
 DB 942 CCGCGCGCACTGGGCGTCACTCCCTTACGTTTCGCTTACCAACCAACGCTACCGGCT 1001  
 QY 2562 GCTGATGAGCGCTCTGCAACCGCGCGCTGCTGCGGCTTCCGCGCACTTACATTGGCG 2621  
 DB 1002 GGTGATGACAGGCGCGGCACTGCGCTGAGCTGGGTGCGGCTGATGATTTGAGTGGCG 1061  
 QY 2622 GGTGCGGCTGCGCACTGGGCTACTGCGCGATGAGAGAGAGAGCGCGCAAGAGCTTCTGCT 2681  
 DB 1062 CATTTGGGCTCGCGAGGCTATTTTCAACGATCCCTGCTGCGAGCAGCAATTTTAC 1121  
 QY 2682 GCACCCCGAGACCGGCGGCGCTCTTACAGACCGCGCATGTGGGCTGCGCTCCGCGCA 2741  
 DB 1122 GCTCCCGA-----CGAGCGCTGTATCGACCGCGCATCTCGGCTGCTACTGGCGAGA 1175  
 QY 2742 TGAACATTCAGATTTCATGGGCGGTGAGAGACAACTCAAGTTCAGCTTCCGATACCGGCT 2801  
 DB 1176 TGGACATTCAGATTCTCGGTGCTGCGGACAGAGGTCAAGTTCGAGATATTCGAT 1235  
 QY 2802 TGAGCTGGGGAATTCAGAGAACTCAATGCGATCGATCGAAGCGAGCGCGGTAT 2861  
 DB 1236 CGAGCTGGCGAATTCAGAAACCGGCTTACGACGCTGCGGCGGTGAACAGCAACCT 1295  
 QY 2862 TGTGCGGCTCGG 2873  
 DB 1296 TCTGGCATCGG 1307

#### RESULT 12

AAFB1367  
 ID AAFB1367 standard; DNA; 7335 BP.

XX AAFB1367;

XX 04-JUN-2001 (first entry)

XX Quorum sensing controlled gene gsc109 ORF.

XX Quorum sensing; antibacterial; bacterial signalling;

XX opportunistic pathogen; immunocompromised; burn; cystic fibrosis;

XX immunosuppressive therapy; AIDS; ss.

XX *Pseudomonas aeruginosa*.

XX W0200118248-A2.

XX 15-MAR-2001.

XX 01-SEP-2000; 2000WO-US24141.

XX 03-SEP-1999; 99US-0153022.

XX (IOWA) UNIV IOWA RES FOUND.

XX (QUORUM-) QUORUM SCI INC.

XX Whiteley M, Lee KM, Greenberg EP, Muhl U;

XX WPI; 2001-265973/27.

XX Identifying modulators of quorum sensing signaling in *Pseudomonas*

PT aeruginosa bacteria, useful for treating infections in  
PT immunocompromized patients -  
xy

Claim 46; Page 108-110; 115pp; English.

Bacteria signal to one another to coordinate expression of specific genes in a cell density dependent fashion. This "bacterial signaling" is called "quorum sensing and response". Quorum sensing allows a bacterial species to sense its own number and regulate gene expression according to population density. The present sequence is an open reading frame (ORF) of a *Pseudomonas aeruginosa* quorum sensing controlled gene. Inhibitors of quorum sensing signaling renders a bacterial population more susceptible to treatment. The present invention relates to a method for identifying modulators of quorum sensing signaling in *Pseudomonas aeruginosa* bacteria. Modulators of quorum signaling may be used to treat *P. aeruginosa* infections. *P. aeruginosa* is an opportunistic pathogen of immunocompromised individuals (burn patients, cystic fibrosis patients, patients undergoing immunosuppressive therapy and patients with AIDS).

Sequence 7335 BP; 1184 A; 2256 C; 2625 G; 1270 T; 0 other;

Query Match	6.7%;	Score	284.2;	DB	22;	Length	7335;
Best Local Similarity	53.4%;	Pred	No. 9.1e-51;				
Matches	698;	Conservative	0;	Mismatches	588;	Indels	21;
						Gaps	4

Qy	1574	TGACACGGCTCTTTCGGCGCGCGGCTCGACGACAGCTCCCATCAGCTCCCTCGTGTCTCG	1633
Qy	1574	TGACACGGCTCTTTCGGCGCGCGGCTCGACGACAGCTCCCATCAGCTCCCTCGTGTCTCG	1633
Db	1004	TGCAACCGGTTGTTCGAGGAGCAGGCTCGAGCGCAACGCCGACCGCGCGCTGGCTTCG	1063
Qy	1634	CGCGCAGACGCTCACGTCACGAAGAGCTTTGGCCCTGTTCCGGCGGAGACTTTGGCGCGGAC	1693
Db	1064	GCGAGGAACGCTTGACTACACCCGACGCTGAACCGCGCGGGCCCAACCGCTTCGCGCATGCC	1123
Qy	1694	TGCGCGAGACAGGGGGACGCCCGGAACATTTGGTCGCGGTGATGTGAGAAAGGCTGGG	1753
Db	1124	TGATTCGACGGCGGGGTTCGGTCCGACCGCCCTGTGTGGGGTGTGGCCATGGAAGGCTTCATCG	1183
Qy	1754	AGCAGGTTGTTCGGGTTCTCGGGGTGCTCGAGTCAGGCGCGGCTTACGTGCCATTCGATG	1813
Db	1184	AGATGTGTCGTGGCCCTGATGCGATCTTCGAAGCGCGGGCGGCTCACTCGCCGGGTGAGCC	1243
Qy	1814	CCGACCTTACCGGCGGAGGCTATCCACATTCACCTTCGTGATCATGTGAGGTAAAGCTCGAG	1873
Db	1244	CGGAGTACCCCGAGAGCGCCAGGCTCACTCATCTGTGAAGACAGCGCGGTGACGCTGCTCG	1303
Qy	1874	TGACGACAGCCATGCGCTGAGTGGCAAACTGTATGCGCCGCGCGGGATTCAGCGGCTGCTCG	1933
Db	1304	TCAGCAGATCGCACCTTGAAAGCTGCGCGCTGGCGCAAGCGGTGACGAGATCGACCTTGACC	1363
Qy	1934	TGACCGAAGCGCGGCGCTCGAAGGCCAGCGGCGACAGCTTCGATGATGCGCCATTCAAGAAC	1993
Db	1364	AGGCGGATGCTTGCTGGAAAAACATGCGGAGAACATATCGGGAGT---CGAGCTGAACG	1420
Qy	1994	CTTGGGATCTCGCGTATGTCATCTTACACTCGGGGATTCACAGGGTTGGCCCAAGGGGGTGA	2053
Db	1421	GCGAGAACTTGTGCTATGATCTTACACTCTCGGCTCCACCGGAGCAAGGAGTGGCG	1480
Qy	2054	TGATTCATCATCGGGGTCCCGTCAACACATCTTGACATCAAGAGCGCTTCGAATAG	2113
Db	1481	GCAACCGGCAATTCGGCGCTGAGCAACCGCTTTGTGTGATGCAGACGGCTTACCGGCTTGG	1540
Qy	2114	GAGCCGAGACAGGGGTGCTGGGGCTCTCTCGCTGAGTGTGGATCTCTCGGCTTATGATG	2173
Db	1541	GCGTCGCGACACGAGTGTTCAGAAAGACCCGTTTCAGCTTTGACGTGTGGATCTTGGAGT	1600
Qy	2174	TGTTTCGGGATCTTCGCGGCGGGCGGTACGATCGTGGTGCAGACGCGTCCAAGCTGCGCG	2233
Db	1601	TCTTCTGCGCTGATGATGGGGGACGTTTGTGTGTGGCGCGCGCGGTGACATTCGCG	1660
Qy	2234	ATCCGCGCCATTTGGCACAATTTGATTCGAACGAGAGAAAGTGAAGGTGTGGAACTTCGTGTC	2293
Db	1661	ACCGCGCAGAGCTGTGTGCGCTGATCAACCGGAAGGGGTGCACACGCTGCATCTTCGTGTC	1720

Oy	2294	CGGGCGTGAATGCGGATGCTGTCAGACATTTTGAAGGGTCTCCCCGATTCGCTGCTAGT	2353
Db	1721	CGTGAATGCTGACAGCCTTCTCTGAGAGACGAAGACGTGCTCTCTGACCAAGC-----	1773
Oy	2354	CTCTGCGGCTTTGCTGCTGCTAGCGGGCAGCTGATCCCGTGAGCTCTCGTGCGAGCTCC	2413
Db	1774	--CTGAAGACGATGTTGGACGCGGCGAGGCGCTGTGCGGAGCGCCACAGACGAGGTGT	1831
Oy	2414	AGGCGATCAGGCCCGGGGTGTGTCGGTGAATCAAGCCGTGGGGGGGCGCACCGAAGCGTCGATCT	2473
Db	1832	TCGCGAAGCTGCGCCGACAGCGCGGCTCTATAACCTCTATAGCGCCGACCGAAGCGGCAT--	1889
Oy	2474	GGTGCATCGGGTACCCGCTGAGGAACGTGCACCTATGCTGGGCGAGCATCCCTCAACGACC	2533
Db	1890	----CGACGTACCCCATCTGAGAGCTGCTGTGAGAGAGGCGACAGACGCGGTCCATTCGGCC	1945
Oy	2534	GTCGCGTGCAGMACAGACGTTCCACGCTGTGATGAGCGCTGAAACCGCGCCGCTCT	2593
Db	1946	GGCCGATGCGCAACTGGGCTGTCAATCTCGATGCGACCTGGAACCGGTCGCGGTGG	2005
Oy	2594	GGGTTCCGGGGCACTCTACATTGAGCGGGGTTCGGGCTGCGACATGAGGCTACTGGGCGCATG	2653
Db	2006	GCGGTCTCGGAGACTGTACTCTGGCCGCTCGGGGCTCGGCTCGTGAGGTACACACAGGCTC	2065
Oy	2654	AAGAGAGACGCGCAAGAGCTTCTCTGTGACCC---CGAGAACGGGGAGCGGCTCTACA	2710
Db	2066	CGGGGCTGACTGCCGAGCGTTTCTGTCCGACGCCCTTCTGTGCTGAGGAGCGAGATGTACC	2125
Oy	2711	AGACCGGCGGATCTTGGGCGCGCTACCTGTGCCGATGAGAACATCGAGTTCAATGGGCGTGAAG	2770
Db	2126	GCAACGGGAGACTGTGGCGGGCTACCGGCGCCGATGGGAGTGAAGTACGACCGCGGGCGGATCG	2185
Oy	2771	ACAAACCAATCAAGCTTTCGGGATACCGCGTTGAGCTTCGGGAAATGAGAGAAACGTTCA	2830
Db	2186	ACCAACAGGTGAACCTGCGCGGCTCTGCGCATTCGAGCTGGGCGAATGAGGCGCGCTGC	2245
Oy	2831	AGTGCATTCGCAACGTACGCGAGCGGATGATGCGCCGTTCGGGAC	2877
Db	2246	TGAGACATCCGTGGGTGCGAGGCGCGGCTGCTGCGGCTGACACGC	2292

RESULT 13	
AA11992	AA11992 standard; DNA; 37856 BP.
XX	
AC	AA11992;
XX	
DT	07-AUG-2000 (first entry)
DE	S. cellulosum DNA encoding polyketide and hereropolyketide enzymes.
XX	
XX	Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis
KW	epithione; cytotoxic; immunosuppressant; antibiotic; antifungal;
KW	plant-protection; ds.
XX	
OS	Sorangium cellulosum.
XX	
XX	Location/Qualifiers
FH	complement (3398..6100)
FT	/tag= a
FT	/product= "ORF1-cRNA synthetase"
FT	/note="gtg start codon"
FT	complement (6374..7111)
FT	/tag= b
FT	/product= "ORF2-monooxygenase"
FT	complement (8433..9550)
FT	/tag= c
FT	/product= "ORF3-aminotransferase"
FT	/note="AGT start codon given in the specification"
FT	9855..11393
FT	/tag= d
FT	/product= "ORF4- tyrosine/DOPA-Decarboxylase"
FT	/note="GTG start codon"
CDS	

```
FT CDS 12212..13658
FT /tag= e
FT /product= "ORF5-3-oxoacyl-ACP-reductase"
FT /note= "ACC start codon"
FT CDS 15374..19984
FT /tag= f
FT /product= "ORF6-polyketide synthase"
FT CDS 20003..27889
FT /tag= g
FT /product= "ORF7-peptide synthetase"
FT CDS 28251..29400
FT /tag= h
FT /product= "ORF8-transpeptidase"
FT CDS complement (30040..31720)
FT /tag= i
FT /product= "ORF9-regulation element"
FT /note= "CGC stop codon"
FT CDS 31982..32932
FT /tag= j
FT /product= "ORF10-transcription regulator"
FT CDS 33128..33613
FT /tag= k
FT /product= "ORF11-regulation element"
FT /note= "GTG start codon"
FT CDS 33661..34077
FT /tag= l
FT /product= "ORF12-regulation element"
FT CDS complement (35255..35616)
FT /tag= m
FT /product= "ORF13-transcription regulator"
FT CDS complement (35730..36242)
FT /tag= n
FT /product= "ORF14-transcription regulator"
FT /note= "GTG start codon"
XX DE19846493-A1.
XX 13-APR-2000.
XX PD
XX 09-OCT-1998; 98DE-1046493.
XX PF
XX 09-OCT-1998; 98DE-1046493.
XX PR
XX (GBFB ) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX Beyer S, Mueller R;
XX WPI: 2000-294101/26.
XX DR
XX DNA sequence coding for products involved in the biosynthesis of
XX polyketide or heteropolyketide compounds, especially epothilone
XX Claim 3; Page 20-33; 36pp; German.
XX
XX This invention describes a novel DNA sequence (I) whose expression
XX products effect or are involved in the enzymatic biosynthesis,
XX mutasynthesis or partial synthesis of polyketide or heteropolyketide
XX compounds (II). (I) can be inserted into an expression vector and used
XX to transform or transfect prokaryotic or eukaryotic cells with the aim
XX of obtaining strains that produce large amounts of polyketide or
XX heteropolyketide compounds, especially epothilones, which have cytotoxic
XX and/or immunosuppressant and antibiotic and antifungal activities and
XX are useful as plant-protection agents. This sequence represents the DNA
XX sequence isolated from Sorangium cellulosum which is described in the
XX method of the invention.
XX
XX Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;
SQ
Query Match 6.3%; Score 265; DB 21; Length 37856;
Best Local Similarity 46.7%; Pred No. 1.6e-46;
Matches 1219; Conservative 0; Mismatches 1325; Indels 69; Gaps 9;
```

```
Db 20259 TGCAGTATGACCCCTCGCTGCGCATCGAGTGCCTCGTGTGAGACCTGAGCGGAGATCG 20318
Qy 461 ACCGAGACACAGGGAGAGCGAGCTGTGTGTGTCAGAGATGTCAGACCGCATCT 520
Db 20319 ACGAGCCGAGAGAGAGCGAGAGCGCGGCTCTATGTCGAGAGAGAGCGCGCCCT 20378
Qy 521 ATGACACCGAGCGGCTTCCTCTATGACGTGTGCGCGTTCGCTGAGACGAGCGGCAAA 580
Db 20379 TCGATCTACGCGAGAGCGGCTGTGCTCGAGGCAAGCTGCTCGGTCGCGGAGCGATC 20438
Qy 581 CCGCTCTGTGCTAGTATGATCTCATTAAGTTAGCTTAGGACGCTGTTCATCT 640
Db 20439 ACGTCTGATCTTGAAGATGATCATATGCTCTCGAGCGGTGTGATGAGATGCTGT 20498
Qy 641 TCAAGA-----CTGGCTAGCTTCTAAGAAATCCGAGACTCTCTCCCTG 688
Db 20499 TCAAGAGCTTTCCACGCTTACGCGGCTTTCACAGAGGCGCGCGCTGCGCTCCG 20558
Qy 689 TCTGAGCTCTCGTACCGCGATTATGTACTCGCGCTGAGTCTGCAAGAATCTGAG 748
Db 20559 AGCTGCGATTCATATAGCGGACTTCGCGGTGTGAGCGGAGCTGCTCAGAGCGGAG 20618
Qy 749 CGCATCAACGATCGATGATTACTGGAAGCGGCGCATGCGGAGCTCCACCTCCGCGA 808
Db 20619 TTCTGGAATCGACCTCGGATACGTAGAGAGCACTCCGCGCGCCCAAGCTGTG 20678
Qy 809 CGCTTCGATGAAGCGGATCATCTACCTGGAAGAGATCCGCTTCGCGGACACGAGC 868
Db 20679 AGCTTCGATGAAGCGGATCATCTACCTGGAAGAGATCCGCTTCGCGGACACGAGC 20738
Qy 869 AATGCTGCTCGTGGACTCTGAGGATGATTAAGAGCGGCGTGTGAGGAGCGGAGCTGA 928
Db 20739 TCCGACTCCACTCTCCCTGGAACAGAGCGGTGAGAGCGCTCAGCCGAGAGAGCGGGA 20798
Qy 929 CCGGAGCGGCGTATCTGTGCTGATTTTCCAGATGATCGGCGCTGAGCGGAGCC 988
Db 20799 CCGCTTCATAGACCTGCTGACGCGCTTACAGCGTGTGCTCTCGCTTATGCGGCGGAGA 20858
Qy 989 CCGGTTTACGCTCAACATTAAGCTTTCAACCGGCTCCCGTCCATCCGCGCGTGAAG 1048
Db 20859 GCGATCTGTGTGTGACAGCGCCATTCGGAATTCGACCCGAGAGAG-----TGAAG 20912
Qy 1049 ATATACCGGAGACTTACATGATGATGCTCTCTGATGATGACACCACTCGGCAAGA 1108
Db 20913 GGTGATGCTGCTTCTTGTCAACATCTGTGCGCTGTGATGATCGGGGCGACCCGA 20972
Qy 1109 GCTTGAACAGCGGCTTAAAGCTTCAAGAGCAGCTGTGGAAGCATGATCATCTGCG 1168
Db 20973 GCTTCCGAGACTCTCGGCGGCTGCGGAGGTGACGTTGGGCGCTTACCGGACAGG 21032
Qy 1169 ACGTAAAGGATGAGAGTCCAGGAGAGCGCGCGCGCTCTGGGATTCACAGAGCGG 1228
Db 21033 ACGTCCCTTGAACGCTGTGTGAGAGAGCTGTACCAAGGCGGAGCCCGACCAACCC 21092
Qy 1229 CATTTTCCCGTGTGCTACAGCGCGCTTAAACGAGCATGCG-----TTGGTGTCA 1282
Db 21093 CTTTCTTCAAGTGTCTTCACTGTCAGAAACCCCGATGATGATGATGATGATGATGATG 21152
Qy 1283 CTTGTTGACAGGCTTGAATCTCGGTGTAACACGACGACGACGATCTTCACTGTGTC 1342
Db 21153 ACATTGATCGGGTGGCGCGCGCTGTGGTGAATGAAGCGGCGGAAATTCATGTGATTC 21212
Qy 1343 TGAATCATAGCTTACAGAGACGATGAGGACCTGTCTGTGCGGTGGGACATCG---TCG 1399
Db 21213 TGAAGCTCTCGGAATCGCGGACAGAGGTGCTCGGACGTTGAGATCAACACGACCTGT 21272
Qy 1400 ACGAGTGTTCGCGCGGACCTTCTGAGACGATGTCGAGCGTACGTCGTTTCTTC 1459
Db 21273 TCGAGCGCGGACCATGAGGAGATGCGCGGCTCAGAGTGTGTCTTCCAGCGCG 21332
Qy 1460 GCGGCTCATGAGAAACATAGGAGTAAAGTGGGTAAAGTGGCTGTTCGCTTCGCCAGC 1519
```



Db 21333 TCGCGCGCGCGGATCGACCCATTGCGAGCTGCCCTCATGGGGCCGAGAGCGGAGTC 21392  
Qy 1520 TAGAAGCGCGCGGAGCGCAAAACGCGACCAACGCGTGTGTAAGCGATACGCTGCAG 1579  
Db 21393 GGGATATGGTGGAGTGAATCCACTGCGCGCGCTGTATCCCGAGACCAATTGATCAGC 21452  
Qy 1580 GCCTGTGCGCGCGCGCGGTGAGCACTCCCATGAGCTCCGCGGTGTGTGCGCGCGCA 1639  
Db 21453 AGCTGTTCAGACGACGAGTGAAGCGGTCCCGAGGCGACCGCGGTCTCTCCACGAC 21512  
Qy 1640 AGACGCTCAGTACGAGAGCTTTCCGCGGTTTCGCGGCACTTGGCGCGCGCTCGCG 1699  
Db 21513 ABAAGTTAGATGATGAGAGCTGAACATGGCGCCAAATCAGCTCGGCAATCAGCTGGGA 21572  
Qy 1700 AGCAGGGGCGACGCCGAAACATGTTGTGCGGTGTGATGAGAAAGCTGGAGACAG 1759  
Db 21573 GCCTGGGCGTGGGCCCGAGAGTGCCTGGGTTGTATCTCGAACGCTCAATGAGACGG 21632  
Qy 1760 TTGTGCGGTTCTCGCGGTGCTCGAGTCAAGCGCGGCTCAAGTGGCGATGAGCGGAC 1819  
Db 21633 TGTGCGGATCTCTCGCGGTGCTCAAGCTGAGCGGCGCTCAAGTGGCGATGAGCGT 21692  
Qy 1820 TACCGCGGAGCGTATCACTACCTCTCGATCATGATGAGTAAAGCTGCTGACCG 1879  
Db 21693 ACCCGACGAGCGCTCGGGCTCATATGGCGGACGAGCGCGCTGCTGCTCAGCG 21752  
Qy 1880 AGCCATGCTGATGAGCAAACTGTCAATGGCGCGCGGATCAGCGGCTGCTGTAAG 1939  
Db 21753 AGCGGCTGCTCTCTGAGAGCTGCCCGCCACGCGGATGCAAGCTGTGATCAGCTGAG 21812  
Qy 1940 AGCGCGCGTTCGAGAGCGGCGAC---CAGCTCGATGATGCCATTGACAGACCTT 1996  
Db 21813 CGTGCACAGAGCGCTCTCAGCTGCCACACCTACCCCGAGCGGCGTACCGCC 21872  
Qy 1997 CGGATCTCGGATATGATCATCACTCGGATCAAGGTTGCCAAGGGGTGATGA 2056  
Db 21873 AGAACCTGCATATGATGATCACTCGGCTGAGCGGCGGCGCCAGAGGGGTGCTG 21932  
Qy 2057 TGGATATGCGGGTGCCTGACACATCTGAGCATCAACGAGCGCTTGCAGATGAGG 2116  
Db 21933 TCGAGCAGCGCGGCTCTGACACCTGCCACCGTGAAGCGCAAGCTGTATGAGATGCC 21992  
Qy 2117 CGGAGACAGGGGTGCGGCTCTCTGCTGAGCTTCAATCTGCGGTATGATGAT 2176  
Db 21993 CGGCGACAGGCTCTCTCAGTTCGCGCTGCTCTGCTGACACATGTTCTGCGAATG 22052  
Qy 2177 TCGGATCTCGGCGCGGCGGTACATGCTGTGCGGACGCGTCAAGCTGCGGATC 2236  
Db 22053 CGTGCCTGTTGCTCTCGGGAGGAGCGCTGATGAGGACGCGGAGAGCTTCTCCGG 22112  
Qy 2237 CGGCGATTGGGCGAGTGTGATGAACGAGAGAGGTGACGCTGTGAACTCGTGC 2296  
Db 22113 GACCTCGGCTGCTGAGCTGCTGAAGACACCGGCTCAAGGAGTCTCTGCGCCCTA 22172  
Qy 2297 CGGATGCGGATGCTGCTGAGCATTTTGAAGGTGCGCCCATTCGCTGCTAGGCTC 2356  
Db 22173 CGGTCTCGGAGCGCTGCGCAAGACAGAGCGC-----GCGCTTCCCG 22217  
Qy 2357 TGGGCTTTCGCTGCTGAGCGGCGATGATCCCGGTGAGCGCTGCTGCGAGCTCAG 2416  
Db 22218 TGGGCTGCTGCTGAGTGGCGGAGGAGCGGCGCGGAGCT-----GTCAGACGCT 22271  
Qy 2417 CCATCAGGCGCGCGCTGCTGATCAGCTGGGCGGCGCGACCGAGAGCTGATGAT 2476  
Db 22272 GGAAGGACCGCGAGCGCGCTGTTAACTCTATGCGCGAGACCGAGACCATTTGG 22331  
Qy 2477 CCATCGGATACCGCGTGAAGAGCTGACCTATCGTGGGAGACATCCCTTACGCGCT 2536  
Db 22332 CAAAGCTCCGAGCGGACCTGTCCGAGAAC-----GATCCCGCCATGAGCGCTG 22382  
Qy 2537 CGCTGCGCAACGAGAGCTTCAAGCTGCTGATGAGCGCTGCAACCGCGCGGCTGAG 2596  
Db 22383 CGATTCCCAATATGCAAAATTAATGCTGTGAGAGAGCGCTGAGCGCGGTCCATCGGG 22442

Qy 2597 TTCCGGGCAACTTACATTTGGCGGGGTGCGGCTGGCACTGGGCTTACTGGCCATGAA 2656  
Db 22443 TCCCGGCGAGATCTTCAATCGCGCGGTGGGGGTGCGCCGGGATATCACGGGCGTCCG 22502  
Qy 2657 AGAAGCGCGAAGAGCTTCTCTGTCACCCC-----GAGACCGGGAGCGCTCT 2707  
Db 22503 ACCTAGCGCGGAGGATTTCTATCCCGACCTTGGGGCAACCAAGGGGCGCCCTGT 22552  
Qy 2708 ACAAGCGCGGATTTGGCGGCTTACCTGCGGATGAAACATCGATTATGAGGCGCTG 2767  
Db 22563 ATCGGACCGGATCGGGCGCTGCTGCGGACGGAACCTCAAGTTTCTGCTCGAA 22622  
Qy 2768 AGGACACCAATCAAGCTTCCGGATACCGGTTGAGCTCGGGGAAATCGAGAAAGC 2827  
Db 22623 AGGACGAGAGGTGAAGGTCCGGGTGTCGATGAGCTGAGAGATCCGCGGCGT 22682  
Qy 2828 TCAAGTGCATCCGAACGTAACGACGCGGTGATGTCCTCGGAAACGACGCGCA 2887  
Db 22683 TCTCAAGCACCGCGCGGTGCTGCTCAAGCCGTGCGGTGCGGACAGACGCGGCG 22742  
Qy 2888 ACAAGCTCTTCTTACCTATGTTGTCGCGGAGGCAACGAGACGCGCTGCGAGCAG 2947  
Db 22743 ACAAGCGGCTCGTCCGATATGCTGCGACGCGGAGAGCGCGCTGACCGCGGAGC 22802  
Qy 2948 AGCGAGCTCAAGACGAGCGGATCGAGCGA 2980  
Db 22803 TCGCGAGTCTGAGAGCGAGCATTTGCTGCA 22835

## RESULT 14

AAS1470

ID AAS1470 standard, DNA; 13029 BP.

AC AAS1470;

DT 13-FEB-2002 (first entry)

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #5.

KW Antisense; ds; prokaryotic cellular proliferation gene;

KW antibiotic; antibacterial; drug design.

OS Pseudomonas aeruginosa.

PN WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2001; 2000US-191078P.

PR 23-MAY-2001; 2000US-206848P.

PR 26-MAY-2001; 2000US-207127P.

PR 23-OCT-2001; 2000US-242578P.

PR 27-NOV-2001; 2000US-253625P.

PR 22-DEC-2001; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,

PI Yamamoto RT, Xu HH;

DR WPI; 2001-611495/70.

XX P-PSDB; A033611.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 4052; 511bp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 13029 BP; 1978 A; 4677 C; 4485 G; 1889 T; 0 other;

Query Match 6.2%; Score 261.4; DB 23; Length 13029;

Best Local Similarity 51.8%; Pred. No. 7.5e-46;

Matches 704; Conservative 0; Mismatches 631; Indels 24; Gaps 4;

QY 1572 GCTGACGAGCTGTTCGCGCGCGGGGTGACGAGCTGCCATGACGCTGCGTGTGTC 1631  
 DB 3387 GCTGCGGAGCTGTTCGCGCGCGGGGTGACGAGCTGCCATGACGCTGCGTGTGTC 3446  
 QY 1632 GCGCGGCAAGACGCTACGATGAGAGCTTTGCGCGGTGCGCGGCACTTGCGCGCG 1691  
 DB 3447 GACGCGGCGAGCTTCGCTACGCGCACTGACGCGCGCGGCAACCGCTGCGCGCACTA 3506  
 QY 1692 GCTGCGGAGCG 1751  
 DB 3507 CTTGCGGCAAGAGCG 3566  
 QY 1752 GAGAGAGGTGTCGCGGTTCCTGCGGTGCTGAGTCAGGCGCGCGCGCGCGCGCGCG 1811  
 DB 3567 GCAATGCTGTGCGGTTCCTGCGGTGCTGAGTCAGGCGCGCGCGCGCGCGCGCGCG 3626  
 QY 1812 TGCCGACCTACCG 1871  
 DB 3627 CCCGACCTACCG 3686  
 QY 1872 GCTGACGAGCG 1931  
 DB 3687 ACTGACCGAGCG 3746  
 QY 1932 CGTGACGAGCG 1991  
 DB 3747 CTTGACGAGCG 3806  
 QY 1992 ACCCTGAGCTTCGCGGTTCCTGCGGTGCTGAGTCAGGCGCGCGCGCGCGCGCGCG 2051  
 DB 3807 CGACA---ATCTGCGCTACGATCTACACCTCGCGGTTCACCGCGCGCGCGCGCGCG 3863  
 QY 2052 GATGATGATCATCGCGGTTCCTGCGGTTCCTGCGGTTCCTGCGGTTCCTGCGGTTC 2111  
 DB 3864 GGGCAACACCG 3923  
 QY 2112 AGGGCCGAGACAGAGGTTCGCGGTTCCTGCGGTGCTGAGTCAGGCGCGCGCGCGCGCG 2171  
 DB 3924 CGACGCGCGAGCGAGGTTCGAGTACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3983  
 QY 2172 TGTGTTGAGGATCTGCG 2231  
 DB 3984 GTGCTTCTGCGCGGTTCGAGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTC 4043  
 QY 2232 CGATCCGCGCGATTTGGCGAGATTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2291  
 DB 4044 CGATCCG 4103

QY 2292 GCCGCGCTGATGCGAGATGCTGTCGACATTTTGAAGGTGCGCCCGATTCGCTGCTAG 2351  
 DB 4104 GCCGCGCTGCTGCAATGTTTCATC-----GACGACCGCGCGCGCGCGCGCGCG 4154  
 QY 2352 GTCTCTGCGGTTCGCTGCTGAGCGCGGACTGATATCCCGGTGGGCTGCTGCGGACT 2411  
 DB 4155 CAGCTTGGCGCGCGCTGTTCTCCGCGCGGAGCGCGCTTCGCGGAGCTGCGCAACCGCGT 4214  
 QY 2412 CCAGGCAATGAGCG 2471  
 DB 4215 GTTGCAACGCTTCG 4274  
 QY 2472 CTGTGCTACGCGGTACCCCGTGAAGAGAGTGCAGCTTATGCGGAGCATCCCTACCG 2531  
 DB 4275 -----CAAGTCAACCATTTGGCAGTGCAGCGCGCGGAGACGGCGAGCGTTGCGCGATCGG 4328  
 QY 2532 CCGTCCGCTGGCGCAACGAGCGTTTCAGCTGCTGATGAGCGCTTCGAAACCGCGCGCGT 2591  
 DB 4329 CCGACCGCTGGCGCAACGAGTGTCTGCGCGCTGAGACCGCGAGTTCAACTGTGCGCGG 4388  
 QY 2592 CTGGGTTCCGCGCGCAACTCTACATTTGGCGGCGTGGCGACTGAGCTACGCGCGCGA 2651  
 DB 4389 CGGCTGCG 4448  
 QY 2652 TGAGAGAGAGAGCGCGCAAGAGCTTCTTCG-----TGACCCCGAGACCGCGGAGCGCGT 2705  
 DB 4449 CCCGCGCTGAGCT 4508  
 QY 2706 CTACAGAGCGCGCGCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2765  
 DB 4509 GTACCG 4568  
 QY 2766 TGAGAGAGAGAGCGCGCAAGAGCTTCTTCG-----TGACCCCGAGACCGCGGAGCGCGT 2825  
 DB 4569 CTTGACCGAGAGGCTGAGCTGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4628  
 QY 2826 GCTCAAGTGCATCGAGAGCTGACGCGAGCGCGGTATGTCGCGCGCGCGCGCGCGCG 2885  
 DB 4629 CTTGCG 4688  
 QY 2886 GACCAAGCTCTTCTGAGCTATGTCGTCGCGGAGCGCGC 2924  
 DB 4689 CAGCGAGTGTGCTGAGCTACACCGCGCGCTGTGCGCGC 4727

RESULT 15  
 ABQ78872  
 ID ABQ78872 standard; DNA; 90600 BP.  
 XX  
 AC ABQ78872;  
 XX  
 DT 23-OCT-2002 (first entry)  
 XX  
 DE S. roseosporus daptomycin biosynthetic gene cluster 90kb region.  
 XX  
 XX Daptomycin biosynthetic gene cluster; thioesterase; antibacterial;  
 KW fungicide; virucide; antiparasitic; immunomodulator; antileptic;  
 KW cytotoxic; gene therapy; antimitotic; immunomodulatory; siderophore;  
 KW anti-cholesterolemic; agrochemical; gene; ds.  
 OS  
 XX Streptomyces roseosporus.  
 XX  
 PN WO200259322-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PF 17-OCT-2001; 2001WO-US32354.  
 XX  
 PR 17-OCT-2000; 2000US-240879P.  
 PR 28-FEB-2001; 2001US-272207P.  
 PR 06-AUG-2001; 2001US-310385P.  
 XX  
 PA (MIAO/) MIAO V P W.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:41 ; Search time 15142.8 seconds

(without alignments)  
11435.853 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_11872\_16104  
Perfect score: 4233  
Sequence: 1 atgacgacatcagctctc.....aggcgagagagagctaa 4233

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_hcg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pac:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_srs:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pac:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_srs:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hcg\_hum:\*
- 31: em\_hcg\_inv:\*
- 32: em\_hcg\_other:\*
- 33: em\_hcg\_mus:\*
- 34: em\_hcg\_pln:\*
- 35: em\_hcg\_rtd:\*
- 36: em\_hcg\_mam:\*
- 37: em\_hcg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hngo\_hum:\*
- 40: em\_hngo\_mus:\*
- 41: em\_hngo\_other:\*

score greater than or equal to the score of the result being printed,  
and its derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4233	100.0	68750	1	AF210843
2	4233	100.0	68750	1	AF193029
3	4233	100.0	68750	6	AR199551
4	4233	100.0	68750	6	AR199559
5	4233	100.0	68750	6	AR199567
6	4233	100.0	68750	6	AR201097
7	4233	100.0	68750	6	AR208671
8	4114.6	97.2	58733	1	AF217189
9	4114.6	97.2	71989	6	AR172664
10	944.8	22.3	42603	1	AF188287
11	728	17.2	77457	1	AF210249
12	634.6	15.0	189050	1	AL646066
13	609.4	14.4	40156	1	AF516145
14	600	14.2	302007	1	SC0939132
15	592.6	14.0	189050	1	AL646066
16	577.6	13.6	18023	6	AX024285
17	550.6	13.0	311143	1	AE016762
18	543.4	12.8	4305	6	AX276331
19	543.4	12.8	12303	1	AE009297
20	543.4	12.8	31220	1	AE008316
21	543.4	12.8	48715	6	AX276312
22	543.4	12.8	50538	6	AX702471
23	538.8	12.7	5430	1	AF184622
24	526	12.4	17071	1	AE004839
25	508.4	12.0	15023	1	AF074705
26	476.8	11.3	290850	1	SC0939127
27	454.4	10.7	4317	1	AF184621
28	406.6	9.6	309904	1	AE016865
29	395	9.3	7444	1	AE013843
30	395	9.3	35806	1	AF091251
31	395	9.3	119443	1	YP102KB
32	395	9.3	311143	1	AE016762
33	395	9.3	335050	1	AL414150
34	390.4	9.2	32159	6	AX702468
35	383.6	9.1	6468	1	YEP1RP2H
36	360.6	8.5	135638	1	AF484556
37	358.2	8.5	10570	1	MTCY22H8
38	358.2	8.5	19782	1	AE007084
39	358.2	8.5	306550	1	EX248342
40	333.8	7.9	41156	1	MSGY222
41	329.2	7.8	276800	1	SC0939115
42	317.4	7.5	216050	1	AL646076
43	296.8	7.0	23451	1	AB070952
44	296.8	7.0	302300	1	AP005034
45	287.2	6.8	15719	1	AE004667

#### ALIGNMENTS

RESULT 1  
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000  
DEFINITION Sorangium cellulosum strain So ce90 epoch1ione biosynthesis gene cluster, complete sequence.  
ACCESSION AF210843  
VERSION AF210843.1 GI:6724237  
KEYWORDS  
SOURCE  
ORGANISM Polyangium cellulosum  
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
Sorangineae; Polyangiaceae; Polyangium.  
REFERENCE  
1 (bases 1 to 68750)  
Mojnar, I., Schupp, T., Ono, M., Zitzke, R., Milnamow, M.,  
Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D. D.,

Pred. No. is the number of results predicted by chance to have a

TITLE  
Goriach, J., Mayo, J.M., Hu, A., Goff, S., Schmid, J. and Ligon, J.M.  
The biosynthetic gene cluster for the microtubule-stabilizing  
agents epothilones A and B from *Sorangium cellulosum* So ce90  
JOURNAL  
Chem. Biol. 7 (2), 97-109 (2000)  
MEDLINE  
20130945  
PUBMED  
10662695  
REFERENCE  
2 (bases 1 to 68750)  
AUTHORS  
Molnar, I.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-DEC-1999) Natural Product Genetics, Novartis  
Agribusiness Research Institute, Inc., 3054 Cornwallie Rd, P.O.Box  
12257, Research Triangle Park, NC 27709, USA

FEATURES  
source  
Location/Qualifiers  
1..68750

/organism="Polysangium cellulosum"  
/mol\_type="genomic DNA"  
/strain="So ce90"  
/db\_xref="taxon:56"  
<1..1826  
/note="ORF1"  
/codon\_start=3  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26904.1"  
/db\_xref="GI:6724238"

## CDS

/translation="ASLDALFARATSARVLDGKATERRHYLAERGIEDRLAREH  
LRIQEGSPFHCMLGDLIVELLADQPLASISFPHARSLRHPDWTSDMLVDGELV  
RLAARGAGPLREYEEERERARQAQEARLMLAAPCFAPDLRFEDDAGLPLGP  
MSPEVAERRLRASVATPELCAALAMIGKQGWGSPYAVEM,PEMLILGFGIPT  
ATAAASPGTSEALRGARLFPASMEVSSKSGCIGNIPEALMBERLRTIVRAMGNAN  
LSRPEREIAAEVRLRQAPAPAGAGLAAGVSSGRLGIVTDGALYSGGND  
LYMFORGISPVVLACTDPFELAPLSOML,FAHANAQTTISKUTCESPILIVARN  
QARPSLVHARGFMAVNOAMVDPDEGAPFVQRTIMEFHPPTPCHEPAGSAFS  
LACDEHLWCELSAGRLMRPHRPPAPSPFALIGHPIAATWYPSLTINATHVL  
MADPBRALIGVDRKGTVEPIVLAETRHPPAHVSEDRDIPALTGQPSDRMWHVHIR  
SGASTVADYQOLMDRPMVMNRLPFTMDRILTLARS"  
complement(1900..3171)  
/note="ORF2: similar to hypothetical proteins from  
Mycobacterium tuberculosis and Streptomyces coelicolor and  
to Dp-peptidases"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26905.1"  
/db\_xref="GI:6724239"

## CDS

/translation="MGALISVAAPGCALGAEEOGPQDAGAGALAPAREVMAEVA  
ACQMPGAVMLVARGDVDVDAVGTSLGSAAPRRDITPRIASMTKAVATAVMMLVE  
EGKLDLDSVDRMLPELAPARKVLARIDGPIDETPAREPIVADLMFTMGEGISDA  
SSPIORAIIDELGVNAQVPMTPHGPDWIRRLGTLPLMHOGAQMVMVTGSLQGV  
VGRANDQGFADATYRERILAPLGMRODPHYVPADKLARFAGCCGFTDEQCEKTRMD  
GASAYASPAPFPGAGLAVSTDVDTLTPARKLMNGVHGRLLSASARENTADHL  
TQAOKAASSFPQGFETGWMYGMAVVTAPDVAVEVGRYGMDDGFGTSMINDPGL  
IGIVMTQSGAFLFSGALREFRMSVYVATESA"  
3415..5556  
/note="ORF3: similar to Na/H and drug/H antiporters"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26906.1"  
/db\_xref="GI:6724240"

## CDS

/translation="MHGILTEROVLISVLTALITVAPASGELARLROPEVILGELFG  
GVYLGPSVYGALAPGFHRALFQEPAYGVVLSCISWYCALLLIMAGIEDVGLIKREA  
RFGALSALGAIAPPLAAGAFSALVLDRLPSGLFGLVLSVAVSIVAKVILERSM  
RSYAVQTLAAGVSEVAAVMVLAMTSSSGSPALAVASLALSGFLFMVLVGR  
LTHLMRWADATRVSKQGVSLVLTFLAALITOTGLPILGALFALLVLSNAPRT  
NRPILDVOTLVAGLPAPFVLVAGRWVDSOLRTPAAGTVALLATLTAAXVAPAA  
LCARLGGLEKSPALVAVGLNKKGRDILVAVIGVGLISNEAYTMVAVVALLTYTA  
SPALLIMTKRAPPTQESARLEREBAARAYTPQVRLIVPIVNAALGPRATDYES  
IVASKRKIGETVDTLTVSQOAPGPSRAGASRGLAGRLVGRVGRORRELTGS  
IQAILASRDHDLVLVGARSPARAGMSFRLQDAIVQRAESVILVVDPPAPAES  
ARRIVPIIGLEYSFAADLAHALAVALMADELTRETLARAPYDLVLGCVYDHQPLGRL  
YLGSTYESVYVNSRVALLVHAGTREQVR"

## CDS

complement(5612..5992)  
/note="ORF4"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26907.1"  
/db\_xref="GI:6724241"  
/translation="MKDPIGRTRCAAGYIPGSGNPEPOMTSHETACLNASORDA  
OVALIVYSDDEPAGPARVTPARTTRHVRNDLTPEPPIPRDDIYASVIESDAIIV  
QHTRLDSQAENALSTIAYTDRE"  
6226..6675  
/note="ORF5"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26908.1"  
/db\_xref="GI:6724242"

## CDS

/translation="MKHYDGTGRFRFRIRIGHTGLLASMALACGSPSEKTVQGTPLAP  
GADARVTDVDPDATTRLVADVHLSPERLEAGSERFVVMORSPSPSPRRVGVLD  
YNADSRKRLKLAETTPYANFELLITAEKQSSPQSPSSAAVIGTSPVC"  
7610..11875  
/gene="epoa"  
7610..11875  
/gene="epoa"  
/note="EPoS A"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAF26919.1"  
/db\_xref="GI:6724253"

## CDS

/translation="MADRPIRAAEDPIAIVGASGRLPGVYIDLSGFMTLLGSDTV  
GRVAPERDAAAMPDPDPDARGKTPVTRASTLSYDACPDAFPGISPREALRMDAHR  
LLEECWALEENALABSAVGETGVGTIGPSEYEAALQOATSAETIDAGCGTGM  
PSVQGRISYALGKPCVADVAATSSLVVHACQSLKELSGARADGRLAVIRG  
STLWGLSTRBALGKAGPAADEAGDEGREGCAVVLKELSGARADGRLAVIRG  
SAINHDGASGLTVNGSSOETVLKRALADGCAASVYGEAHGTGTLGPIEIOA  
LNAYIGARDVATPLLIGSVKTNLGHPEYASGITLILVLSLOHGOI,PAHLHAOLN  
PRISGDI,RLVYTRAKTPPMMNTPRACVSSFCMSSGNNAHVVEEAPACTCPAPE  
RPAELIVASRTASALDQAKLRDHLSTYSCIGDAPFASLATTTRSMERHLVAVAT  
SREGRLAALDAAAOQOTS,PGAVRSIADSSRKLAVLFTGQAQTLGKMGRLYDVSAF  
REAPDLCTVRLFNQOELRPLREVMAPASVDAALDQTAFTOPALFTEGVALLAMRS  
MGVEPELVAGHSIGELVAVAGVAGSLDAVFLVARGRLMOALPAGGAMYSIEAPEA  
DYAAVAVHAAASVSIAVNAVDQVYIAGGQPVHAIAMARAGARTALAVSHAFS  
PLMAPLEAFGRVAESVSRPSIYLVNLSGKACTDVSRRGVRARAREVRADG  
VKAHAGAGTFVEVGPSTLGLVPAKCPMPDARPLLASSRAGREPRVYLEALGIM  
AVGGLVSWAGLFPSSGRRVPLPTYPWQERRWITKADADGDRAGAGADEVEEG  
GAVRGDGRSSARLDHPPEPSGRKEVAGDRPFRLIDEQGVLDHLVARTERAPG  
LGEVIAVADAAGLSENDQALAGVPPDDLPGKPNPILGSCAGRIYAVGCVNGLV  
VGQPIYALASAGFAHVTTSALVLPFOALSAIPAAAMPVAYLTAAVLDRIALQV  
GERVYIHAATGVGLAAVMOHVGAEVHATAGPEKRAYLESGVRVYVDSRDPFV  
ADVRAWTGSEGVYVNLNLSGELIDKSNILRSRHRFVLEKRDCAVADNOGLPRLR  
NLSFSLVLRKGMERPARVRLLEBELGLLAAGVFTPPPIATLTIARVADAFRMAQ  
AOHLQKLVLTLDDEVOQIRIPHAAGAGSTDRDLDDLSAAPARAALAEAFRTQ  
VSOVLRTEIKVGAELFTRLGMDSLMAGVDRNRLEASLKTSLTFLSTSPNALLA  
ONLDALATLSLIERVAENLRAGVQNDVFSAGDQDWEIIAL"  
7640..8920  
/gene="epoa"  
/note="Region: beta-ketoacyl synthase/malonyl-CoA  
decarboxylase"  
9236..10201  
/gene="epoa"  
/note="Region: acetate acyltransferase"  
10529..11428  
/gene="epoa"  
/note="Region: enoyl reductase"  
11549..11764  
/gene="epoa"  
/note="Region: acyl carrier protein"  
11872..16104  
/gene="epop"  
11872..16104  
/gene="epop"  
/note="EPoS P: contains condensation/heterocyclization,

## CDS

cysteine adenylation, putative NAD(P)H oxidase, and  
 peptidyl carrier domains  
 /codon\_start=1  
 /transl\_table=11  
 /product="nonribosomal peptide synthetase"  
 /protein\_id="AAF26925.1"  
 /db\_xref="gi:6724259"  
 /translation="MTINQLNLELHQIKLADGERLQAPKNAIPNLARISH  
 KSTILMLRPLASIVPAPERRAPPLTDIOESYLGRTAFVSGHAREYD  
 CTDLDPRLSPAFKRVARHMDLRAHLPDMQVILEPVADIDILRGLDRSTRA  
 RLVSRLDMSRIRYDTERPLXHYVAIVDRQRLYSIDLINVDLSISITKDWL  
 SFYDPEPSELVLELSTRDYVLALESRKSAHQSDMYKRLIAELPEPTLWKAD  
 PSTLKEIRFRTEQMLPSDSWGLRKRRGEGRLPTGVILAAFEVIGRMASRFTL  
 NITLENRLPHRVNDITGDTSMWLDITTRDKSPFOAKRLOEOLMEMDCHDS  
 GIEVREARVIGIORGALFVPLVLSALNDQOVVTSIQRGTPVYSTPOLLHD  
 QLYEHDDVLVAMDIQGVPPDLDMLEAVVPLRLREVEGEVORCSLPPAOLE  
 ARASANTALSHETLHGLFPAVVEOLPMLAVASARKITTYEELSRSRLCARLR  
 EOGAPNTLVAVMEKMEVQVAVLVESGAVYPIADIPAKRIHLHDGKVLV  
 LTQPMIDGKLSWPPGIQRLVSEAGVEDGQPPMPPIQTSVDLAVIYTSGLPK  
 GVMIDHKGAVNTIIDINERFEIGPDVLTALSLSFDLSVDVGIILAGGTIVPDA  
 SKLRDPAHMAELIEREKTVMNSVPALMRMLVEHEGEPDLSLASRLSLSGMIPV  
 GLPELQIRGVVISLIGATASISWIGVPRVNDLSMASITVGRPLRQTFHVD  
 EALERPVMVPGQLYIGVGLALGYMRDEETKRSFLVHPETGRUXKXGGLYLD  
 GNIEFWGEDNQILKRGVLELGETLTKSHPVNDVAVIPVANDANLKLLAVVLP  
 EGTRRAAEODASLTERIDARAHAAEDLSDERVOFKLARERDLDEKVVVLP  
 TGQDPREAGLDVVARRSVTFLEAPIFVEFRFLSLSSVEPDGATLPFRYPSAG  
 STYVOTAAVYKGRLEGVDEGFYHYFHEHRLKLSHGERGHANRONVDDEAA  
 FNLLFVRIDIESLSSSPRECLLEGAOALMRBACNIGVCPQFNEQVR  
 PVLDRHSDVYHGMLAGRVDPPOFVCTTLQDSSPRATTRGAPRGREOFADMLRD  
 FLRTKLPEYVIVFVELDALPLTNSKGVDKALREKDTSPHSGDANLBEI  
 LVAVREVLGLVEVGLQOSFVLDGATSIIHVRMSLLQKRLREIATTELFOYNLDS  
 LASGLRDSRLDQPMQDRVEVRKRRRS"  
 16251..21749  
 /gene="epob"  
 16251..21749  
 /note="Epos B"  
 /codon\_start=1  
 /transl\_table=11  
 /product="polyketide synthase"  
 /protein\_id="AAF26920.1"

Query Match 100.0%; Score 4233; DB 1; Length 68750;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGACGATCAATCAGCTTCTGAAGAGCTCGAGCACCGGATATCAAGCTGGCGCGCAT	60
DB	11872	ATGACGATCAATCAGCTTCTGAAGAGCTCGAGCACCGGATATCAAGCTGGCGCGCAT	11931
QY	61	GGGAGGCGCTCCAGATCAGGCCCCCAAGAACGCCCGAACCAGATCTGCTCGTGA	120
DB	11932	GGGAGGCGCTCCAGATCAGGCCCCCAAGAACGCCCGAACCAGATCTGCTCGTGA	11991
QY	121	ATCTCCGAGCAAAAGACGATCTCGAGATGCTCGTCAAGAGATCCCGCGAGATCC	180
DB	11992	ATCTCCGAGCAAAAGACGATCTCGAGATGCTCGTCAAGAGATCCCGCGAGATCC	12051
QY	181	ATCTGTCGCGCCCGCAGCGGAGCGGACGCTCGTTCTCTCAAGACATCCAGATCC	240
DB	12052	ATCTGTCGCGCCCGCAGCGGAGCGGACGCTCGTTCTCTCAAGACATCCAGATCC	12111
QY	241	TACTGGCTGGGCGGAGACGCTTACGGTCCCAAGGAGATCCAGCGCTATCGCGAA	300
DB	12112	TACTGGCTGGGCGGAGACGCTTACGGTCCCAAGGAGATCCAGCGCTATCGCGAA	12171
QY	301	TACGACTGTACGATCTCGAGCTGCGAGGCTGAGCCGCGCTTTCGAAAGTCGTCCG	360
DB	12172	TACGACTGTACGATCTCGAGCTGCGAGGCTGAGCCGCGCTTTCGAAAGTCGTCCG	12231
QY	361	CGGACGACATGCTTTCGGGCGCACGCTGCCCGACATGATGACGATGAGCTTAA	420
DB	12232	CGGACGACATGCTTTCGGGCGCACGCTGCCCGACATGATGACGATGAGCTTAA	12291

QY	421	GTCCAGCCCGACATCGAGATCATGATCTGCGCGGGCTCGACCGGACACACGGGAAGCG	480
DB	12292	GTCCAGCCCGACATCGAGATCATGATCTGCGCGGGCTCGACCGGACACACGGGAAGCG	12351
QY	481	AGGCTCGTGTCTTTCGAGATGCGATGTCGACCGCATCTATGACACCGAGCCCTCCG	540
DB	12352	AGGCTCGTGTCTTTCGAGATGCGATGTCGACCGCATCTATGACACCGAGCCCTCCG	12411
QY	541	CTCATACGCTCGCGCGCTGCGCTGGAAGAGCGGAAACCGCTCGTCTGATGATC	600
DB	12412	CTCATACGCTCGCGCGCTGCGCTGGAAGAGCGGAAACCGCTCGTCTGATGATC	12471
QY	601	GATCTCATTAACGTTTGAAGAGAGCGCTGTCATCTTCAAGACCTGCTCAGCTTC	660
DB	12472	GATCTCATTAACGTTTGAAGAGAGCGCTGTCATCTTCAAGACCTGCTCAGCTTC	12531
QY	661	TACGAGATCCCGAGACCTCTCTCTCTGCTGGAAGCTCTGATCCGCGATTATGATC	720
DB	12532	TACGAGATCCCGAGACCTCTCTCTCTGCTGGAAGCTCTGATCCGCGATTATGATC	12591
QY	721	GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATGAGATGATGATTAATGAGCGG	780
DB	12592	GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATGAGATGATGATTAATGAGCGG	12651
QY	781	CGCATCCCGAGCTCCGACCTCCGCGAGCGCTTCCGATGGAAGCCGATCCATACCTG	840
DB	12652	CGCATCCCGAGCTCCGACCTCCGCGAGCGCTTCCGATGGAAGCCGATCCATACCTG	12711
QY	841	AAGGAGATCCGCTTCCGCGACACGAGCAATGCTCCGCTGGAAGCTCTCGGGTGCATG	900
DB	12712	AAGGAGATCCGCTTCCGCGACACGAGCAATGCTCCGCTGGAAGCTCTCGGGTGCATG	12771
QY	901	AAGCGGCTGTCCGCGAGCGCGGCTGACCCCGAGCGGCTGATCTGCGGCAATTTCC	960
DB	12772	AAGCGGCTGTCCGCGAGCGCGGCTGACCCCGAGCGGCTGATCTGCGGCAATTTCC	12831
QY	961	GAGGTGATCCGCGCTGGAAGCGCGAGCCCGGCTTACGCTCAACATTAACGCTTTCAAC	1020
DB	12832	GAGGTGATCCGCGCTGGAAGCGCGAGCCCGGCTTACGCTCAACATTAACGCTTTCAAC	12891
QY	1021	CGGCTCCCGCTCATCCGCGGCTGAACGATATCAACCGGAGATTACGTCGATGCTCTC	1080
DB	12892	CGGCTCCCGCTCATCCGCGGCTGAACGATATCAACCGGAGATTACGTCGATGCTCTC	12951
QY	1081	CTGACATCCACACCACTCCGCGACAGAGCTTCAAGCGCGGTAGCGTATCAAGAG	1140
DB	12952	CTGACATCCACACCACTCCGCGACAGAGCTTCAAGCGCGGTAGCGTATCAAGAG	13011
QY	1141	CAGCTGAGGAAGGATGATCACTGCGACGTAAGCGTATCGAGTCCAGCGAGAGCC	1200
DB	13012	CAGCTGAGGAAGGATGATCACTGCGACGTAAGCGTATCGAGTCCAGCGAGAGCC	13071
QY	1201	GCCCGGCTCTGGGGAATCCACGAGCGCATTTGTTCCCGTGATGCTCAAGAGCGGCTT	1260
DB	13072	GCCCGGCTCTGGGGAATCCACGAGCGCATTTGTTCCCGTGATGCTCAAGAGCGGCTT	13131
QY	1261	AACGAGCAATGCTTGGTGTCACTGTTGCAAGAGGCTCCGGAATCCCGGTATACCAAGC	1320
DB	13132	AACGAGCAATGCTTGGTGTCACTGTTGCAAGAGGCTCCGGAATCCCGGTATACCAAGC	13191
QY	1321	ACGAGACTCTCAGCTGCTGCTGATCATCAGCTTACAGAGCAGATGGGAGCCTCGTC	1380
DB	13192	ACGAGACTCTCAGCTGCTGCTGATCATCAGCTTACAGAGCAGATGGGAGCCTCGTC	13251
QY	1381	CTCGCTGGGACATGCTGCAAGAGTGTCCCGCCGACCTTCTGAGCAGATCTCGAA	1440
DB	13252	CTCGCTGGGACATGCTGCAAGAGTGTCCCGCCGACCTTCTGAGCAGATCTCGAA	13311
QY	1441	GCGTACGCTGTTTTTCTCCGCGGCTCACTGAGGAACCATGGGGTGAACAGTGCCTGT	1500
DB	13312	GCGTACGCTGTTTTTCTCCGCGGCTCACTGAGGAACCATGGGGTGAACAGTGCCTGT	13371

QY 1501 TCGCTTCGCGCTGCGCCAGCTAGAGCGCGGCGCAAGCGCAACGCGCTGCTG 1560  
Db 13372 TCGCTTCGCGCTGCGCCAGCTAGAGCGCGGCGCAAGCGCAACGCGCTGCTG 13431  
QY 1561 AGGAGCATACGCTGCAAGCGCTGTTTCGCGCGCGGGTCGAGCAGCTGCGCATGACCTC 1620  
Db 13432 AGGAGCATACGCTGCAAGCGCTGTTTCGCGCGCGGGTCGAGCAGCTGCGCATGACCTC 13491  
QY 1621 GCGGTGTGTGCGCGCGCAAGAGCTCAGTACGAAAGAGCTTTTCGCGCGCTGCGCGCA 1680  
Db 13492 GCGGTGTGTGCGCGCGCAAGAGCTCAGTACGAAAGAGCTTTTCGCGCGCTGCGCGCA 13551  
QY 1681 CTTCGCGCGCGCTGCGCGCAAGAGCGCGCGCAAGCTTTTCGCGCGCTGCGCGCA 1740  
Db 13552 CTTCGCGCGCGCTGCGCGCAAGAGCGCGCGCAAGCTTTTCGCGCGCTGCGCGCA 13611  
QY 1741 GAGAAAGCTGAGAGAGGTTTTCGCGCTTTCGCGGTGCTGAGTCAAGCGCGCGCTAC 1800  
Db 13612 GAGAAAGCTGAGAGAGGTTTTCGCGCTTTCGCGGTGCTGAGTCAAGCGCGCGCTAC 13671  
QY 1801 GTGCGCATGATGCGCGCACTACCGCGCGAGCTGATCACTCTCTCGATCATGCTGAG 1860  
Db 13672 GTGCGCATGATGCGCGCACTACCGCGCGAGCTGATCACTCTCTCGATCATGCTGAG 13731  
QY 1861 GTAAAGCTGCTGCTGAGCGAGCGATGAGCTGAGGAACTGTCATGCGCGCGCGGATC 1920  
Db 13732 GTAAAGCTGCTGCTGAGCGAGCGATGAGCTGAGGAACTGTCATGCGCGCGCGGATC 13791  
QY 1921 CAGCGCTGCTGCTGAGCGAGCGCTGAGGAGCGAGCGAGCGAGCTTCCGATGATG 1980  
Db 13792 CAGCGCTGCTGCTGAGCGAGCGCTGAGGAGCGAGCGAGCGAGCTTCCGATGATG 13851  
QY 1981 CCCATTCAACACACTTCGCGATCTCGCGTATGATCACTCACTCCGCGATTCACAGGTTG 2040  
Db 13852 CCCATTCAACACACTTCGCGATCTCGCGTATGATCACTCACTCCGCGATTCACAGGTTG 13911  
QY 2041 CCCAAGGGGTGATGATCATCATCGAGGTCGCGTCAACACATCTCTGACATCAAGAG 2100  
Db 13912 CCCAAGGGGTGATGATCATCATCGAGGTCGCGTCAACACATCTCTGACATCAAGAG 13971  
QY 2101 CCGTTGAAATAGAGGCGCGGAGACAGAGTGTGCGCTCTCTCGTGAAGCTTCGATCTC 2160  
Db 13972 CCGTTGAAATAGAGGCGCGGAGACAGAGTGTGCGCTCTCTCGTGAAGCTTCGATCTC 14031  
QY 2161 TCGGTATGATGATGTTTCGCGATCTCTGCGCGCGGCGGATACATCTGTGTGCGGACCG 2220  
Db 14032 TCGGTATGATGATGTTTCGCGATCTCTGCGCGCGGCGGATACATCTGTGTGCGGACCG 14091  
QY 2221 TCCAAAGCTGCGCATCCGCGCATTTGGGAGAGTTGATCGAAGAGAGAGAGTGAAGCGTG 2280  
Db 14092 TCCAAAGCTGCGCATCCGCGCATTTGGGAGAGTTGATCGAAGAGAGAGAGTGAAGCGTG 14151  
QY 2281 TGGAACTGCGTGCAGGCGCTGATGCGAGTCTCTGAGCACTTTTGAAGGTCGCCCAT 2340  
Db 14152 TGGAACTGCGTGCAGGCGCTGATGCGAGTCTCTGAGCACTTTTGAAGGTCGCCCAT 14211  
QY 2341 TCGCTGCTAGAGTCTTCGCGGCTTTTCGCTGCTGAGAGGCGACCTGATCCCGGTCGCTG 2400  
Db 14212 TCGCTGCTAGAGTCTTCGCGGCTTTTCGCTGCTGAGAGGCGACCTGATCCCGGTCGCTG 14271  
QY 2401 CCGTGCAGAGCTCCAGGCGCATACAGGCGCGGCGTGTGCGTATCGAGTCCGCGCGGCGCAC 2460  
Db 14272 CCGTGCAGAGCTCCAGGCGCATACAGGCGCGGCGTGTGCGTATCGAGTCCGCGCGGCGCAC 14331  
QY 2461 GAAAGCTGATCTGCTCATTCGGGTACCCCGTGAGAGAACGTGACATCTGTGAGGCGAGC 2520  
Db 14332 GAAAGCTGATCTGCTCATTCGGGTACCCCGTGAGAGAACGTGACATCTGTGAGGCGAGC 14391  
QY 2521 ATCCCTTACGCGCGCTCCGCTGCGCAACAGAGTTCAGTGTTCATGAGGCGCTCGAA 2580  
Db 14392 ATCCCTTACGCGCGCTCCGCTGCGCAACAGAGTTCAGTGTTCATGAGGCGCTCGAA 14451  
QY 2581 CCGCGCGCGGTCTGAGGTTCCGCGGCAACTTACATTTGCGCGGCTCGGCGCTGCGAGC 2640

Db 14452 CCGCGCGCGGTCTGAGGTTCCGCGGCAACTTACATTTGCGCGGCTCGGCGCTGCGAGC 14511  
QY 2641 TACTGCGCATGAAAGAGAGCGCGCAAGAGCTTCTCTGTCACATCCCGAGACCGGGAG 2700  
Db 14512 TACTGCGCATGAAAGAGAGCGCGCAAGAGCTTCTCTGTCACATCCCGAGACCGGGAG 14571  
QY 2701 GCGCTCTACAGACCGGCGATGTCGCGCTACCTGCGCGAGTGAAGAACATGAGTTCAATG 2760  
Db 14572 GCGCTCTACAGACCGGCGATGTCGCGCTACCTGCGCGAGTGAAGAACATGAGTTCAATG 14631  
QY 2761 GCGCGTGAAGACCAACCAATCAAGCTTCGCGATACCGGTTGAGCTCGGGAATCGAG 2820  
Db 14632 GCGCGTGAAGACCAACCAATCAAGCTTCGCGATACCGGTTGAGCTCGGGAATCGAG 14691  
QY 2821 GAAAGCTCAAGTGCATTCGACCTGACGCGAGCGCGTGTATGTCCTCGTGGAAACGAC 2880  
Db 14692 GAAAGCTCAAGTGCATTCGACCTGACGCGAGCGCGTGTATGTCCTCGTGGAAACGAC 14751  
QY 2881 GCGCGGAACAGCTCTTACCTATGATGTCGCGAGGCGACACGAGACGCGCTGCGC 2940  
Db 14752 GCGCGGAACAGCTCTTACCTATGATGTCGCGAGGCGACACGAGACGCGCTGCGC 14811  
QY 2941 GAGCAGACGCGAGCTCAAGACCGAGCGGATTCAGCGAGAGCACAGCGCCGCAAGCG 3000  
Db 14812 GAGCAGACGCGAGCTCAAGACCGAGCGGATTCAGCGAGAGCACAGCGCCGCAAGCG 14871  
QY 3001 GACGCTTGAAGCGAGCGGAGGATTCAGGCTTCGCTGACACGAGCTCCGAGAG 3060  
Db 14872 GACGCTTGAAGCGAGCGGAGGATTCAGGCTTCGCTGACACGAGCTCCGAGAG 14931  
QY 3061 GACCTGAGCGGAACCGCTGCTGATGACCGCGGAGAGTCCGCGGAGCGGAGGCTG 3120  
Db 14932 GACCTGAGCGGAACCGCTGCTGATGACCGCGGAGAGTCCGCGGAGCGGAGGCTG 14991  
QY 3121 GACGCTACGCGCGCTGCGGATGCGCTGCGAAGCTTCTTGAAGGCCCGATTCGTTGTT 3180  
Db 14992 GACGCTACGCGCGCTGCGGATGCGCTGCGAAGCTTCTTGAAGGCCCGATTCGTTGTT 15051  
QY 3181 GAGTTTGTGCTATTCCTGAGCTCTTGAAGCGAGCTGAGACCGGCGGACCTTCCC 3240  
Db 15052 GAGTTTGTGCTATTCCTGAGCTCTTGAAGCGAGCTGAGACCGGCGGACCTTCCC 15111  
QY 3241 AAATTCGTTATCCATCCGCGGCGAGCGATACCGGAGCAACCTACCGATGATCAAA 3300  
Db 15112 AAATTCGTTATCCATCCGCGGCGAGCGATACCGGAGCAACCTACCGATGATCAAA 15171  
QY 3301 TCCGCGCGCATGAGGCGGTGACGAGGCTTCTATTATTCACACCGTTCCAGACCGT 3360  
Db 15172 TCCGCGCGCATGAGGCGGTGACGAGGCTTCTATTATTCACACCGTTCCAGACCGT 15231  
QY 3361 TTGCTGAAGCTTCCGATCAAGGATTCAGACCGGAGCGACGCTTCCGCAAACTTCCAG 3420  
Db 15232 TTGCTGAAGCTTCCGATCAAGGATTCAGACCGGAGCGACGCTTCCGCAAACTTCCAG 15291  
QY 3421 GTTTCGATGAAGCGGCGTTCAACCTCCGTTTCGCGGAGGATTCAGACCGATCGAGT 3480  
Db 15292 GTTTCGATGAAGCGGCGTTCAACCTCCGTTTCGCGGAGGATTCAGACCGATCGAGT 15351  
QY 3481 CTGTATGATCGTCTGCGGAGATTTTTCCTGCTGAGGCGGATTTATGCGGAGCTC 3540  
Db 15352 CTGTATGATCGTCTGCGGAGATTTTTCCTGCTGAGGCGGATTTATGCGGAGCTC 15411  
QY 3541 CTGATGAGACAGGCGCTTCTGCAACATCGGCGCTGTCGCGTGGGCAATTCATTTT 3600  
Db 15412 CTGATGAGACAGGCGCTTCTGCAACATCGGCGCTGTCGCGTGGGCAATTCATTTT 15471  
QY 3601 GAAAGGTTTCGCGCGGTTCTGACCTGAGACATTCGAGCTTTTACGTCGAGCGAGCTG 3660  
Db 15472 GAAAGGTTTCGCGCGGTTCTGACCTGAGACATTCGAGCTTTTACGTCGAGCGAGCTG 15531  
QY 3661 GCGCGGCGGTTGAGCCCGCGGATTCAGGCTGATGAGCTTCAGGATTCCTACAGC 3720



Db	15532	GGCGGGGGGGTAAGACCCGGCGGACAGTTCCAGAGTCGTACGCTCGAGTCAGAGATTCTCACGG	15531
QY	3721	AGGGCGCCCAACGACGCGCGGGCGCCCTCCCGCGCGGAGCAGACACTTCGCCGATATGCTTT	3780
Db	15592	AGGGCGCCCAACGACGCGCGGGCGCCCTCCCGCGCGGAGCAGACACTTCGCCGATATGCTTT	15651
QY	3781	CGGCACTTCATTGAAGGACCAAACTAACCCGAGTACATGGGTGCTACAGTCTTCGTGAGGCTC	3840
Db	15652	CGGCACTTCATTGAAGGACCAAACTAACCCGAGTACATGGGTGCTACAGTCTTCGTGAGGCTC	15711
QY	3841	GATCGCTTTGGCCGCTGACGTCCAAACGCGCAAGTGCATGCTGAAGGCCCTGCGCGAGCGGAAG	3900
Db	15712	GATCGCTTTGGCCGCTGACGTCCAAACGCGCAAGTGCATGCTGAAGGCCCTGCGCGAGCGGAAG	15771
QY	3901	GATACCTTCGTGCGCGCGGCAATTCGGGGCACACGCGCCACCGGAGCGCTTGGAGAGATC	3960
Db	15772	GATACCTTCGTGCGCGCGGCAATTCGGGGCACACGCGCCACCGGAGCGCTTGGAGAGATC	15831
QY	3961	CTCGTCGCCGGTCCGTACCGGGAGGTGCTCGGGCTGAGGTGGTCCGGGCTCCAGCAGAGCTTC	4020
Db	15832	CTCGTCGCCGGTCCGTACCGGGAGGTGCTCGGGCTGAGGTGGTCCGGGCTCCAGCAGAGCTTC	15891
QY	4021	GTCGATCTTGGTGGCGACATCGATTCACATGCTTCGCAATGAGAGGCTGTGGCAAGAGAG	4080
Db	15892	GTCGATCTTGGTGGCGACATCGATTCACATGCTTCGCAATGAGAGGCTGTGGCAAGAGAG	15951
QY	4081	CTGGATGAGGAGATCGCCATCACCGAGTTGTCACGATCCCGAACCCTCGGCTCGCTGGCG	4140
Db	15952	CTGGATGAGGAGATCGCCATCACCGAGTTGTCACGATCCCGAACCCTCGGCTCGCTGGCG	16011
QY	4141	TCCGGTTTGGCCCGGAGACTGGAAGATCTAGATCAGCGGCGCGAACAATGCAAGACCGAGTG	4200
Db	16012	TCCGGTTTGGCCCGGAGACTGGAAGATCTAGATCAGCGGCGCGAACAATGCAAGACCGAGTG	16071
QY	4201	GAGGTTGGGGCAAGGCGAGAGACGAGCTAA	4233
Db	16072	GAGGTTGGGGCGCAAGGCGAGAGACGAGCTAA	16104

[illegible]

Db	11992	ATCTCCGAGGACAAAGAGCAGATCTCTGAGGATCTCCGTCCAGAGACTCCCGGCAAGATCC	12051
Qy	181	ATCGTCCCGCCGCCAGCCGAGCGGACGCTCCGTTTCTCTGACAGACATCCAGAAATCC	240
Db	12052	ATCGTCCCGCCGCCAGCCGAGCGGACGCTCCGTTTCTCTGACAGACATCCAGAAATCC	12111
Qy	241	TACTGCTGGGCGCGACAGGAGGCTTTACGGTCTCCGAGGGGATCCAGCCCTATCGCGAA	300
Db	12112	TACTGCTGGGCGCGACAGGAGGCTTTACGGTCTCCGAGGGGATCCAGCCCTATCGCGAA	12171
Qy	301	TACGACTGTACGGATCTCCGACGTCGCGAGGCTGAGCCGGCCCTTTGGAAAGTCGTCGAG	360
Db	12172	TACGACTGTACGGATCTCCGACGTCGCGAGGCTGAGCCGGCCCTTTGGAAAGTCGTCGAG	12231
Qy	361	CGGACGACATGCTTCGGGCGCCACACGCTGCCGACATGATCAGGTGATTCGAGCTTAA	420
Db	12232	CGGACGACATGCTTCGGGCGCCACACGCTGCCGACATGATGAGGTGATTCGAGCTTAA	12291
Qy	421	GTCGACGCGCATTCGAGATCATCGATCTTCGCGGGCTTCGACCGGAGCACCGGAAAGCG	480
Db	12292	GTCGACGCGCATTCGAGATCATCGATCTTCGCGGGCTTCGACCGGAGCACCGGAAAGCG	12351
Qy	481	AGGCTGTGTCTGTGGAGATGGAGATGGGACCGGATCTATGACACGAGGGCCCTCCG	540
Db	12352	AGGCTGTGTCTGTGGAGATGGAGATGGGACCGGATCTATGACACGAGGGCCCTCCG	12411
Qy	541	CTGTATCACGTCCTCCGCTCGGCTGAGACGAGCGGCAACCCGTTCTGTCGTCAAGTATC	600
Db	12412	CTGTATCACGTCCTCCGCTCGGCTGAGACGAGCGGCAACCCGTTCTGTCGTCAAGTATC	12471
Qy	601	GATCTCATTTAACGTTAGACCTTAGGACGCTGTCCATCATCTTCAAGACCTGGCTCAAGCTTC	660
Db	12472	GATCTCATTTAACGTTAGACCTTAGGACGCTGTCCATCATCTTCAAGACCTGGCTCAAGCTTC	12531
Qy	661	TACGAAGATCCCGAGACCTCTCTCCCTGTCTCTGAGCTCTCGTACCGGATTTATGTATCTC	720
Db	12532	TACGAAGATCCCGAGACCTCTCTCCCTGTCTCTGAGCTCTCGTACCGGATTTATGTATCTC	12591
Qy	721	GCCCTGGAAGTCTGTGCAAGAGTCTGAGGGGCAATCAACGATTCGATGGAATTACTGGAAGCGG	780
Db	12592	GCCCTGGAAGTCTGTGCAAGAGTCTGAGGGGCAATCAACGATTCGATGGAATTACTGGAAGCGG	12651
Qy	781	CGCATCGCGGAGCTCCACCTTCGCGCGACGCTTCGAGTAAAGCGGATCATCTACCTCG	840
Db	12652	CGCATCGCGGAGCTCCACCTTCGCGCGACGCTTCGAGTAAAGCGGATCATCTACCTCG	12711
Qy	841	AAGGAAGATCCGCTTCCGGGACACAGAGCAATGGCTTCGGAATCTCTGGGGGTGATTTG	900
Db	12712	AAGGAAGATCCGCTTCCGGGACACAGAGCAATGGCTTCGGAATCTCTGGGGGTGATTTG	12771
Qy	901	AAGCGGCGGTGTGGGGGAGGCGGGGCGTGAACCGGAGCGGAGTATCATCTGGCTGATTTTCC	960
Db	12772	AAGCGGCGGTGTGGGGGAGGCGGGGCGTGAACCGGAGCGGAGTATCATCTGGCTGATTTTCC	12831
Qy	961	GAGGTGATCGGGGCGTGGAGCGGAGGCCCGGTTTACGCTCAAGATTAACGCTCTTCAAC	1020
Db	12832	GAGGTGATCGGGGCGTGGAGCGGAGGCCCGGTTTACGCTCAAGATTAACGCTCTTCAAC	12891
Qy	1021	CGGCTCCCGGTTCATTCGCGCGGTGAACGATATCACCGGGGACTTCAAGTGCATGCTCTC	1080
Db	12892	CGGCTCCCGGTTCATTCGCGCGGTGAACGATATCACCGGGGACTTCAAGTGCATGCTCTC	12951
Qy	1081	CTGGAATCGGACACCATGCGGACAAAGACTTCGAAACGCGGGGCTTAAGGGTATTCAAAGG	1140
Db	12952	CTGGAATCGGACACCATGCGGACAAAGACTTCGAAACGCGGGGCTTAAGGGTATTCAAAGG	13011
Qy	1141	CAGCTGTGGGAAGCGATGTGATCACTGCGACGTAAAGCGGTATGAGGTCCAGCGAGAGCC	1200
Db	13012	CAGCTGTGGGAAGCGATGTGATCACTGCGACGTAAAGCGGTATGAGGTCCAGCGAGAGCC	13071
Qy	1201	GCCCGGCTCTGGGGATCCACGAGGCGCATTTGTTCCCGTGTGCTACGAGCGCGCTT	1260

Db 13072 GCCCGGTCCTGGGATCCAGAGGCGGATTTCTCCCGTGTCTCAAGAGCGGCTT 13131  
Qy 1261 AACCAAGATCGTTGGTGTCACTCTGTTGACAGAGGCTCGAACTCCGGTGTACACAC 1320  
Db 13132 AACCAAGATCGTTGGTGTCACTCTGTTGACAGAGGCTCGAACTCCGGTGTACACAC 13191  
Qy 1321 AGCGAGCTCTCAAGTGTGTGTATCAATACGCTTACAGAGACGATGGGACCTCCGTC 1380  
Db 13192 AGCGAGCTCTCAAGTGTGTGTATCAATACGCTTACAGAGACGATGGGACCTCCGTC 13251  
Qy 1381 CTCGCGTGGAGATCTCGACGAGATGTTCCGCGCCGACCTTCTGAGACGATGCTCGAA 1440  
Db 13252 CTCGCGTGGAGATCTCGACGAGATGTTCCGCGCCGACCTTCTGAGACGATGCTCGAA 13311  
Qy 1441 GCGTACGTGTTTTTCTCCGCGGCTCTACTGAGAAACATGGGGTGAACAAGTCCGCTGT 1500  
Db 13312 GCGTACGTGTTTTTCTCCGCGGCTCTACTGAGAAACATGGGGTGAACAAGTCCGCTGT 13371  
Qy 1501 TGGCTTCCGCTCCGACGTAAGAGCGGGGCGAGGCAAAAGCGACAAAGCGCTGCTG 1560  
Db 13372 TGGCTTCCGCTCCGACGTAAGAGCGGGGCGAGGCAAAAGCGACAAAGCGCTGCTG 13431  
Qy 1561 AGCGAGCATACGCTGACGCGCTGTTTCGCGCGCGGGTGCAGCAGCTGCCATGACGCTC 1620  
Db 13432 AGCGAGCATACGCTGACGCGCTGTTTCGCGCGCGGGTGCAGCAGCTGCCATGACGCTC 13491  
Qy 1621 GCGGTGTGTCCGCGCGCAAGACGCTCAAGTACGAAGAGCTTTCCGCGCTTCCGCGCA 1680  
Db 13492 GCGGTGTGTCCGCGCGCAAGACGCTCAAGTACGAAGAGCTTTCCGCGCTTCCGCGCA 13551  
Qy 1681 CTGGGCGCGCGCTGCGAGAGAGGCGGGGACCGCCGACAACTAGTCCCGGGTGGATG 1740  
Db 13552 CTGGGCGCGCGCTGCGAGAGAGGCGGGGACCGCCGACAACTAGTCCCGGGTGGATG 13611  
Qy 1741 GAGAAAGCTGGGAGAGGTTGTCCGCTTCTCGCGTCTCGAGTCAGAGCGCGGCTTAC 1800  
Db 13612 GAGAAAGCTGGGAGAGGTTGTCCGCTTCTCGCGTCTCGAGTCAGAGCGCGGCTTAC 13671  
Qy 1801 GTGCCGATGATGCGGACCTTACCGCGGAGCGGATATCACTCTCTGATCATGGTGA 1860  
Db 13672 GTGCCGATGATGCGGACCTTACCGCGGAGCGGATATCACTCTCTGATCATGGTGA 13731  
Qy 1861 GTAAAGCTGTGTGACGAGGCTATGCGGATGGGAACTGTCAATGAGCGCGCGGAGATC 1920  
Db 13732 GTAAAGCTGTGTGACGAGGCTATGCGGATGGGAACTGTCAATGAGCGCGCGGAGATC 13791  
Qy 1921 CAGCGGCTCTGTGACGAGGCGCGGCTGTAAGAGCGAGCGGACCGCTCGATGATG 1980  
Db 13792 CAGCGGCTCTGTGACGAGGCGCGGCTGTAAGAGCGAGCGGACCGCTCGATGATG 13851  
Qy 1981 CCCATTCAACACCTTCCGATCTCCGCTATGTCATCTACCTCCGGATTCACAGGTTG 2040  
Db 13852 CCCATTCAACACCTTCCGATCTCCGCTATGTCATCTACCTCCGGATTCACAGGTTG 13911  
Qy 2041 CCCAAGGGGATATGATGATCATCGAGGAGCGGTCACACCACTCTGACATCAAGAG 2100  
Db 13912 CCCAAGGGGATATGATGATCATCGAGGAGCGGTCACACCACTCTGACATCAAGAG 13971  
Qy 2101 CGCTTGAATAGGGCCCGGAGACAGAGTGTGCGGCTCTCTCGTGAAGTTGATCTC 2160  
Db 13972 CGCTTGAATAGGGCCCGGAGACAGAGTGTGCGGCTCTCTCGTGAAGTTGATCTC 14031  
Qy 2161 TCGGTCTATGATGTGTTCCGGAATCTTGCGCGCGGGCGGTAAGATCGTGTGCGGAGCGG 2220  
Db 14032 TCGGTCTATGATGTGTTCCGGAATCTTGCGCGCGGGCGGTAAGATCGTGTGCGGAGCGG 14091  
Qy 2221 TCCAACTCGCGATCCGCGGCAATGGGCGAGGTTATGAAACGAGAGAAAGTGAAGG 2280  
Db 14092 TCCAACTCGCGATCCGCGGCAATGGGCGAGGTTATGAAACGAGAGAAAGTGAAGG 14151  
Qy 2281 TGGAACTCGGTGCGCGGCTGATGCGGATGCTGTGAGCATTTGAGGGTCCGCCGAT 2340  
Db 14152 TGGAACTCGGTGCGCGGCTGATGCGGATGCTGTGAGCATTTGAGGGTCCGCCGAT 14211

Qy 2341 TGGCTCGTAGTCTCTGCGGCTTCTGCTGTGAGCGCGGACTGTGATCCCGTGGGCTG 2400  
Db 14212 TGGCTCGTAGTCTCTGCGGCTTCTGCTGTGAGCGCGGACTGTGATCCCGTGGGCTG 14271  
Qy 2401 CCTGCGAGCTCCAGGACCATCAGGCGCGGCTGTGATGATGAGCTGGGCGGAGC 2460  
Db 14272 CCTGCGAGCTCCAGGACCATCAGGCGCGGCTGTGATGATGAGCTGGGCGGAGC 14331  
Qy 2461 GAAAGCTGATCTGATCATCGGGTAACTCCGTTAGAAACGTGACCTATCTGGGCGAGC 2520  
Db 14332 GAAAGCTGATCTGATCATCGGGTAACTCCGTTAGAAACGTGACCTATCTGGGCGAGC 14391  
Qy 2521 ATCCCTTACGCGCGCTCCGCTGGGCAACGAGCTTCCACGCTGATGAGAGGCGCTCGAA 2580  
Db 14392 ATCCCTTACGCGCGCTCCGCTGGGCAACGAGCTTCCACGCTGATGAGAGGCGCTCGAA 14451  
Qy 2581 CCGCGCGGATCTGGGTTCCGGGCGAACTTACATTTGCGGGGTGGGCTGGCACTGGGC 2640  
Db 14452 CCGCGCGGATCTGGGTTCCGGGCGAACTTACATTTGCGGGGTGGGCTGGCACTGGGC 14511  
Qy 2641 TACTGCGGATGAAAGAGAGAGCGCGAAGCTTCTGTGACCTCCGAGACCGGGAG 2700  
Db 14512 TACTGCGGATGAAAGAGAGAGCGCGAAGCTTCTGTGACCTCCGAGACCGGGAG 14571  
Qy 2701 CGCCTTACAAAGACCGCGGATCTGGGCGGCTTACCTCCGATGGAACATGAGTTTCA 2760  
Db 14572 CGCCTTACAAAGACCGCGGATCTGGGCGGCTTACCTCCGATGGAACATGAGTTTCA 14631  
Qy 2761 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATCCGCTTGAAGCTCGGGAAATCGAG 2820  
Db 14632 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATCCGCTTGAAGCTCGGGAAATCGAG 14691  
Qy 2821 GAAACGTTCAAGTGCATCTCGAACGTACGCGAGCGGCTGATTTGCTCCGTGGGAAACGAC 2880  
Db 14692 GAAACGTTCAAGTGCATCTCGAACGTACGCGAGCGGCTGATTTGCTCCGTGGGAAACGAC 14751  
Qy 2881 GGGGCGAACAAGCTCTTACGCTTATGCTGCTCCGAGGGGCAACAGGAGCGGCTGCG 2940  
Db 14752 GGGGCGAACAAGCTCTTACGCTTATGCTGCTCCGAGGGGCAACAGGAGCGGCTGCG 14811  
Qy 2941 GAGCAGAGCGAGGCTCAAGACCGAGCGGATTCAGCGGAGGACACGCGCCGAAAGCG 3000  
Db 14812 GAGCAGAGCGAGGCTCAAGACCGAGCGGATTCAGCGGAGGACACGCGCCGAAAGCG 14871  
Qy 3001 GACGCGTTGAGCGAGCGGAGGAGGTTCAAGCTTGTGCTGACACGGAATCCGGAAG 3060  
Db 14872 GACGCGTTGAGCGAGCGGAGGAGGTTCAAGCTTGTGCTGACACGGAATCCGGAAG 14931  
Qy 3061 GACCTGAGACGGAAGCGCGTGTGATGATGACCGGGGAGAGTCCGCGGAGGCGGGCTG 3120  
Db 14932 GACCTGAGACGGAAGCGCGTGTGATGATGACCGGGGAGAGTCCGCGGAGGCGGGCTG 14991  
Qy 3121 GACGTCTACGCGCGTGTGAGCTGTGAGCGTTCGGAACGTTCTTGAAGGCCCGATTCGTTGTT 3180  
Db 14992 GACGTCTACGCGCGTGTGAGCTGTGAGCGTTCGGAACGTTCTTGAAGGCCCGATTCGTTGTT 15051  
Qy 3181 GAGTTGTGATCTCTGAGCTGCTTGAAGCAGCGTGAAGCCCGAGCGGACCTTCTCC 3240  
Db 15052 GAGTTGTGATCTCTGAGCTGCTTGAAGCAGCGTGAAGCCCGAGCGGACCTTCTCC 15111  
Qy 3241 AAATTCCGTTATCCATCCGCGGGGAGACGTTACCGGGTCAAAACGCGATGTCAAA 3300  
Db 15112 AAATTCCGTTATCCATCCGCGGGGAGACGTTACCGGGTCAAAACGCGATGTCAAA 15171  
Qy 3301 TCCGCGCGGATGAGAGGCGTGAACGAGGCTTCTATTATTAACACCGTTCCAGACCGT 3360  
Db 15172 TCCGCGCGGATGAGAGGCGTGAACGAGGCTTCTATTATTAACACCGTTCCAGACCGT 15231  
Qy 3361 TTGCTGAAGCTTCTCGATCAAGGATTCAGAGCGGAGCGGACGTTGGGAAACTTTCAGC 3420  
Db 15232 TTGCTGAAGCTTCTCGATCAAGGATTCAGAGCGGAGCGGACGTTGGGAAACTTTCAGC 15291

Qy	3421	GTGTTGCATGAAGGCGGCTTCAACCTCTCTGTGTGGGGAGAGATCGACGCGCATCGAGTGG	3480
Db	15292	GTGTTGCATGAAGGCGGCTTCAACCTCTCTGTGTGGGGAGAGATCGACGCGCATCGAGTGG	15351
Qy	3481	CTGTATGGAATCGTGTGCGAGAAATTTGCTGTGAGGGCCGGATATATATGCGCGAGCTC	3540
Db	15352	CTGTATGGAATCGTGTGCGAGAAATTTGCTGTGAGGGCCGGATATATATGCGCGAGCTC	15411
Qy	3541	CTGTATGAGCAGGGGGCTTCTCTGCAATATGGGGGTCTGTCCGGTGGGGCAATTCATTTT	3600
Db	15412	CTGTATGAGCAGGGGGCTTCTCTGCAATATGGGGGTCTGTCCGGTGGGGCAATTCATTTT	15471
Qy	3601	GAAACAGTTGCGGCGGGTTCTCGACCTTCGACATTTCCGACGTTTACGTCAGCGGACGTG	3660
Db	15472	GAAACAGTTGCGGCGGGTTCTCTGACCTTCGACATTTCCGACGTTTACGTCAGCGGACGTG	15531
Qy	3661	GGCGGGCGGGTAGACCCGCGGAGTTCCAGGTCTGTACGCTCGGTACAGATTCCTCACCG	3720
Db	15532	GGCGGGCGGGTAGACCCGCGGAGTTCCAGGTCTGTACGCTCGGTACAGATTCCTCACCG	15591
Qy	3721	AGGGCGCCCAAGACGGGGGGGCGCCCTCCGGCCGAGAGACATTTGCGGATATGCTT	3780
Db	15592	AGGGCGCCCAAGACGGGGGGGCGCCCTCCGGCCGAGAGACATTTGCGGATATGCTT	15651
Qy	3781	CGCAGCTTCTTTGAGAGCAAACTACCCGAGTACATGGTGGCTTACAGTCTTCGAGAGCTC	3840
Db	15652	CGCAGCTTCTTTGAGAGCAAACTACCCGAGTACATGGTGGCTTACAGTCTTCGAGAGCTC	15711
Qy	3841	GATCGGTTGCCGCTGACGTCCTCAACGGCGAAGTGCATGTATAGGCCCTTGCAGGCGGAAG	3900
Db	15712	GATCGGTTGCCGCTGACGTCCTCAACGGCGAAGTGCATGTATAGGCCCTTGCAGGCGGAAG	15771
Qy	3901	GATACCTCGTCCGCGGGGCAFTTGGGGGCAACGGGCGCAAGGGAAGCTTTGAGAGAGATC	3960
Db	15772	GATACCTCGTCCGCGGGCAFTTGGGGGCAACGGGCGCAAGGGAAGCTTTGAGAGAGATC	15831
Qy	3961	CTCGTCGCGGTCTGTACGGAGGTGCTTGAGGCTGGAGGTGTCTGGGCTCCAGCAGACTTC	4020
Db	15832	CTCGTCGCGGTCTGTACGGAGGTGCTTGAGGCTGGAGGTGTCTGGGCTCCAGCAGACTTC	15891
Qy	4021	GTGCATCTTGGGTGCGACATCGATTACATGTCTTCGATGAGGAGCCTGTTCGAGGAAGG	4080
Db	15892	GTGCATCTTGGGTGCGACATCGATTACATGTCTTCGATGAGGAGCCTGTTCGAGGAAGG	15951
Qy	4081	CTGAGTATGAGGAGATCGCCATCAACCGAGTTGTTCCAGTACCCGAACCTCGGCTGCTGGGG	4140
Db	15952	CTGAGTATGAGGAGATCGCCATCAACCGAGTTGTTCCAGTACCCGAACCTCGGCTGCTGGGG	16011
Qy	4141	TCCGGTTTGGCCGAGACTCGAAGATCTTAGTCAGCGGCCGAACATGACAGACCGAGTG	4200
Db	16012	TCCGGTTTGGCCGAGACTCGAAGATCTTAGTCAGCGGCCGAACATGACAGACCGAGTG	16071
Qy	4201	GAGGTTTCGGCGCAAGGCGCAGAGACCTGTAAGCTAA	4233
Db	16072	GAGGTTTCGGCGCAAGGCGCAGAGACCTGTAAGCTAA	16104

RESULT	3
LOCUS	ARI199551
DEFINITION	ARI199551 68750 bp DNA linear PAT 20-APR-2002
ACCESSION	Sequence 1 from patent US 6355457.
VERSION	ARI199551
KEYWORDS	ARI199551.1 GI:20249625
SOURCE	.
ORGANISM	Unknown.
AUTHORS	Unknown.
TITLE	Unclassified.
JOURNAL	1 (bases 1 to 68750)
FEATURES	Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,T. Genes for the biosynthesis of epothonones Patent: US 6355457-A 1 12-MAR-2002; Location/Qualifiers

source	1. .68750	/organism="unknown"
BASE COUNT	9596 a	22456 c 25539 g 11159 t
ORIGIN		
Query Match	100.0%	Score 4233; DB 6; Length 68750;
Query Local Similarity	100.0%	Pred. No. 0;
Matches 4233; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	ATGACGATCAATCAGCTTCTGAAACGAGCTTGAGCACACAGGGTATCAAGCTGCGCGCAT 60
DB	11872	ATGACGATCAATCAGCTTCTGAAACGAGCTTGAGCACACAGGGTATCAAGCTGCGCGCAT 11931
QY	61	GGGAGGCGCTCCAGATACAGGCGCCCAAGAACGCCCTAACCCGAACTCGCTCGCTGCA 120
DB	11932	GGGAGGCGCTCCAGATACAGGCGCCCAAGAACGCCCTAACCCGAACTCGCTCGCTGCA 11991
QY	121	ATCTCCAGACAAAGACAGATCTGACAGATCTCCGTACAGAGATCTCCCGACAGATCC 180
DB	11992	ATCTCCAGACAAAGACAGATCTGACAGATCTCCGTACAGAGATCTCCCGACAGATCC 12051
QY	181	ATCTGTCGCCCGCCAGCGGACAGCTCGCTCTCTCTACAGACATCCAGATCC 240
DB	12052	ATCTGTCGCCCGCCAGCGGACAGCTCGCTCTCTCTACAGACATCCAGATATC 12111
QY	241	TACTGCGTGGGCGGACAGAGGGTTATGAGTCCCAAGGGAGTCCAGCCCTATGCGGAA 300
DB	12112	TACTGCGTGGGCGGACAGAGGGTTATGAGTCCCAAGGGAGTCCAGCCCTATGCGGAA 12171
QY	301	TACGACTGTACGAGATCTCGACGTGCGGAGCTGAGCGGCGCTTTGCGAAAGTCTGCGG 360
DB	12172	TACGACTGTACGAGATCTCGACGTGCGGAGCTGAGCGGCGCTTTGCGAAAGTCTGCGG 12231
QY	361	CGGACGACATGCTTGGGCGCCACACGCTGCCGACATGATGACAGTATCCAGCTTAA 420
DB	12232	CGGACGACATGCTTGGGCGCCACACGCTGCCGACATGATGACAGTATCCAGCTTAA 12291
QY	421	GTGACGCGCACATTCGAGATCATTCGATCTGCGCGGGCTGACCGGAGCACGCGGAAAGG 480
DB	12292	GTGACGCGCACATTCGAGATCATTCGATCTGCGCGGGCTGACCGGAGCACGCGGAAAGG 12351
QY	481	AGGCTCGTGTGCTTGGCAGATGCGATGTCGACACCGATCTATGACACGAGCGCCCTCGG 540
DB	12352	AGGCTCGTGTGCTTGGCAGATGCGATGTCGACACCGATCTATGACACGAGCGCCCTCGG 12411
QY	541	CTCTATACAGTCTGCGCGCTTGGGCTGAGACGAGCGGCAAAACCGTCTCTGCTCAATATC 600
DB	12412	CTCTATACAGTCTGCGCGCTTGGGCTGAGACGAGCGGCAAAACCGTCTCTGCTCAATATC 12471
QY	601	GATCTCAATTAAAGTTGACCTAGGACAGCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
DB	12472	GATCTCAATTAAAGTTGACCTAGGACAGCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531
QY	661	TACGAGATCCCGAGACCTCTCTCCCTGTCTTGAGACTCTCGTACCGCGATTATGTACTC 720
DB	12532	TACGAGATCCCGAGACCTCTCTCCCTGTCTTGAGACTCTCGTACCGCGATTATGTACTC 12591
QY	721	GCGGTGAGATCTGCGCAAGAGTCTGAGCGGACATCAACGATTCATGATTAATCTGAAAGCGG 780
DB	12592	GCGGTGAGATCTGCGCAAGAGTCTGAGCGGACATCAACGATTCATGATTAATCTGAAAGCGG 12651
QY	781	CGCATCCCGAGCTCCCACTTCGCGCGACGCTTCGATTAAGGCGCATCATCTACCTG 840
DB	12652	CGCATCCCGAGCTCCCACTTCGCGCGACGCTTCGATTAAGGCGCATCATCTACCTG 12711
QY	841	AAGGAGATCCGCTTCCGACACACGAGACATGCTGCGCTCGAGCTCTTGCGGCTCGATTG 900
DB	12712	AAGGAGATCCGCTTCCGACACACGAGACATGCTGCGCTCGAGCTCTTGCGGCTCGATTG 12771
QY	901	AAGCGCGATTCGCGGAGCGCGGAGCTGACCCCGACGCGGCTCATCTGCTGCTGATTTTCC 960
DB	12772	AAGCGCGATTCGCGGAGCGCGGAGCTGACCCCGACGCGGCTCATCTGCTGCTGATTTTCC 12831

OY	961	GAGGTGATCGGGCCCTTGAGAGCGCGAGGCCGCCCGTTTACGTCGAACATAAAGCTTTCAAC	1020
Db	12832	GAGGTGATCGGGCCCTTGAGAGCGCGAGGCCGCCCGTTTACGTCGAACATAAAGCTTTCAAC	12891
OY	1021	CGGCTCCCGGTCATTCGCGCGGTGAAGGATATATCAACGGGAGACTTCAAGTCGATGTCCTC	1080
Db	12892	CGGCTCCCGGTCATTCGCGCGGTGAAGGATATATCAACGGGAGACTTCAAGTCGATGTCCTC	12951
OY	1081	CTGGACATCGACACCACTCGCGACMAAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG	1140
Db	12952	CTGGACATCGACACCACTCGCGACMAAGCTTCGAACAGCGCGCTAAGCGTATTCAAGAG	13011
OY	1141	CAGCTGTGGAGAGGAGATGATATCATCTGGAGAGTAAAGCGGTATCGAGTTCAGCGACAGGCC	1200
Db	13012	CAGCTGTGGAGAGGAGATGATATCATCTGGAGAGTAAAGCGGTATCGAGTTCAGCGACAGGCC	13071
OY	1201	GCCCGGGGTCGAGGGGATATCAACGAGGGGCAATGTTCCCGGTGGGCTCAAGACGGGCTTT	1260
Db	13072	GCCCGGGGTCGAGGGGATATCAACGAGGGGCAATGTTCCCGGTGGGCTCAAGACGGGCTTT	13131
OY	1261	AACGAGCAAGTCGTTGGTGTCACTCTGTCAGAGGCTCGGAATCCGAGTGTACAACACAG	1320
Db	13132	AACGAGCAAGTCGTTGGTGTCACTCTGTCAGAGGCTCGGAATCCGAGTGTACAACACAG	13191
OY	1321	ACGCAAGACTCTCAAGCTGCTGCTGGATATCATAGCTTACAGACACGATGGGGACCTTCCTC	1380
Db	13192	ACGCAAGACTCTCAAGCTGCTGCTGGATATCATAGCTTACAGACACGATGGGGACCTTCCTC	13251
OY	1381	CTCGCGTGGGACATCGTGCAGACGAGGTGTCGCGCGGACCTTCGAGGACGACATGCTCGAA	1440
Db	13252	CTCGCGTGGGACATCGTGCAGACGAGGTGTCGCGCGGACCTTCGAGGACGACATGCTCGAA	13311
OY	1441	GCGTACGTCGTTTTTCTCCGCGGCTCACTGAGGAACATAGGGGTGAACAGTGCCTGT	1500
Db	13312	GCGTACGTCGTTTTTCTCCGCGGCTCACTGAGGAACATAGGGGTGAACAGTGCCTGT	13371
OY	1501	TCGCTTCCGCTGCGCCCAAGCTGAAGCGCGGCGGACGCGCAACCGGCTGCTG	1560
Db	13372	TCGCTTCCGCTGCGCCCAAGCTGAAGCGCGGCGGACGCGCAACCGGCTGCTG	13431
OY	1561	AGCGAGCATACGCTGACACGCGCTGTTGCGGCGCGGGTGAAGAGCGGCCATGACGCTC	1620
Db	13432	AGCGAGCATACGCTGACACGCGCTGTTGCGGCGCGGGTGAAGAGCGGCCATGACGCTC	13491
OY	1621	GCCGTGTGTGTGCGCGCGCCAAAGACGCTCACTGAACGATTTTCGCGCGTTTGCSCGCA	1680
Db	13492	GCCGTGTGTGTGCGCGCGCCAAAGACGCTCACTGAACGATTTTCGCGCGTTTGCSCGCA	13551
OY	1681	CTTGCGCGCGCGGCTGCGCGAGACGAGGGGACACGCCGAACATTTGTCGCGTGTGATG	1740
Db	13552	CTTGCGCGCGCGGCTGCGCGAGACGAGGGGACACGCCGAACATTTGTCGCGTGTGATG	13611
OY	1741	GAGAAAGGCTGGGAGCAGGTTGTTCGCGGTTCTCGAGTCAAGTCAAGCGCGGCGCTAC	1800
Db	13612	GAGAAAGGCTGGGAGCAGGTTGTTCGCGGTTCTCGAGTCAAGTCAAGCGCGGCGCTAC	13671
OY	1801	GTCGCGGATCGATGCGGACCTTACCGCGCGAGCGGATATCATCACTCTCGATCATATGCTGAG	1860
Db	13672	GTCGCGGATCGATGCGGACCTTACCGCGCGAGCGGATATCATCACTCTCGATCATATGCTGAG	13731
OY	1861	GTTAAAGCTCGTGTGACGACGCGCATGCTGATGGAACATGTCATGAGCGCGCGGGGATC	1920
Db	13732	GTTAAAGCTCGTGTGACGACGCGCATGCTGATGGAACATGTCATGAGCGCGCGGGGATC	13791
OY	1921	CAGCGGCTGCTGTGTGACGAGGCGCGGCTGGAAGGCGACCGGCGACCAAGCTCCGATGATG	1980
Db	13792	CAGCGGCTGCTGTGTGACGAGGCGCGGCTGGAAGGCGACCGGCGACCAAGCTCCGATGATG	13851
OY	1981	CCCAATTGAGACACCTTGGGATCTTGCGGTATGTATATCAACCTCGGGATCAACAGGGTTG	2040
Db	13852	CCCAATTGAGACACCTTGGGATCTTGCGGTATGTATATCAACCTCGGGATCAACAGGGTTG	13911
OY	2041	CCCAAGGGGTGATGATCATCGGGGTGCGTCAACCAATCTTGACATCAACGAG	2100

Db	13912	CCCAAGGGGGGTGATGATCATCATCGGGGTGCGGTCAACCACTCTGGACATCAACGAG	13971
OY	2101	CGCTTCGAAATAGAGGCGCCGAGACAAGAGTGTCTGGCGCTCTCGCTGAGCTTCGATCTC	2160
Db	13972	CGCTTCGAAATAGAGGCGCCGAGACAAGAGTGTCTGGCGCTCTCGCTGAGCTTCGATCTC	14031
OY	2161	TCGGTCTATGATGTGTTCCGGGATTCCTGGCGCGCGGGGATACGATCGTGGTCCCGACCGG	2220
Db	14032	TCGGTCTATGATGTGTTCCGGGATTCCTGGCGCGCGGGGATACGATCGTGGTCCCGACCGG	14091
OY	2221	TCCAAGCTGGCGCATCCGGCGCATTTGGGCGAGATTTGATCGAACGAGAGAGGTGACGCTG	2280
Db	14092	TCCAAGCTGGCGCATCCGGCGCATTTGGGCGAGATTTGATCGAACGAGAGAGGTGACGCTG	14151
OY	2281	TGGAACCTCGGTGCGCGCGCTGATCGGATGCTCTGTGAGACATTTTGAGGGGTGCGCCCGAT	2340
Db	14152	TGGAACCTCGGTGCGCGCGCTGATCGGATGCTCTGTGAGACATTTTGAGGGGTGCGCCCGAT	14211
OY	2341	TCGCTCGCTAGTGTCTGCGGGCTTTTCGTGCTGACGCGCGACTGGATCCCGGTGGGCTTG	2400
Db	14212	TCGCTCGCTAGTGTCTGCGGGCTTTTCGTGCTGACGCGCGACTGGATCCCGGTGGGCTTG	14271
OY	2401	CTGGGCGAGCTCCAGGCGCATCAGAGCCCGCGGTGTGGGTATCAGGCTGGAGGCGGGGCGAC	2460
Db	14272	CTGGGCGAGCTCCAGGCGCATCAGAGCCCGCGGTGTGGGTATCAGGCTGGAGGCGGGGCGAC	14331
OY	2461	GAAGCTGTGATCTGTGTTCATCGGGTATCCCGTGGAGAGACGTGCACTTATCGTGGGCGAC	2520
Db	14332	GAAGCTGTGATCTGTGTTCATCGGGTATCCCGTGGAGAGACGTGCACTTATCGTGGGCGAC	14391
OY	2521	ATCCCTTAAGGCGCGTCCGCTGCGCAACAGACCTTCCACGTGTGATGAGGCGCTCGAA	2580
Db	14392	ATCCCTTAAGGCGCGTCCGCTGCGCAACAGACCTTCCACGTGTGATGAGGCGCTCGAA	14451
OY	2581	CCGGGCGCGGTCTGGGTTCCGGGGGCAACTTATATTGGGGGGTGGGGCTGGCACTGGGCT	2640
Db	14452	CCGGGCGCGGTCTGGGTTCCGGGGGCAACTTATATTGGGGGGTGGGGCTGGCACTGGGCT	14511
OY	2641	TACTGGGCGCATGAGAAGAGAGACGCCCAAGACCTTCTCGTGCACCCGAGACCCGGGAG	2700
Db	14512	TACTGGGCGCATGAGAAGAGAGACGCCCAAGACCTTCTCGTGCACCCGAGACCCGGGAG	14571
OY	2701	CGCTCTACAAGACCGGCGATCTGGGCGGCTACCTGCGCGATGGAATACTGATTCATG	2760
Db	14572	CGCTCTACAAGACCGGCGATCTGGGCGGCTACCTGCGCGATGGAATACTGATTCATG	14631
OY	2761	GGGCTGATGAGCAACCAATCAAGCTTCCCGGATACCGGGTGGAGGCTCGGGGAAATCCGAG	2820
Db	14632	GGGCTGATGAGCAACCAATCAAGCTTCCCGGATACCGGGTGGAGGCTCGGGGAAATCCGAG	14691
OY	2821	GAACCGCTCAAGTGTGATCCGAACGTACGCGACGCGGTGATTTGTGCTCCGTCGGGAAACGAC	2880
Db	14692	GAACCGCTCAAGTGTGATCCGAACGTACGCGACGCGGTGATTTGTGCTCCGTCGGGAAACGAC	14751
OY	2881	GCGGCGAAACAGCTCCTTCTAGCGCTATGTGTGTCCGAGAGGCAACAGAGACGCGCTGCTC	2940
Db	14752	GCGGCGAAACAGCTCCTTCTAGCGCTATGTGTGTCCGAGAGGCAACAGAGACGCGCTGCTC	14811
OY	2941	GAGCGAGACCGGAGCTTCMAAGACCGAGCGGATGTGAAGCGAGAGCAACGCGCGCGAAAGG	3000
Db	14812	GAGCGAGACCGGAGCTTCMAAGACCGAGCGGATGTGAAGCGAGAGCAACGCGCGCGAAAGG	14871
OY	3001	GAGCGCTTGAAGCGAGCGCGAGAGGATGACAGTCTCAACTCGCTGCAACACGACTCCGAGAG	3060
Db	14872	GAGCGCTTGAAGCGAGCGCGAGAGGATGACAGTCTCAACTCGCTGCAACACGACTCCGAGAG	14931
OY	3061	GACCTGAGACGAAAGCCCGCTCGTGCATTTGACCGGACAGAGATCCGCGGAGAGCGGGGCTG	3120
Db	14932	GACCTGAGACGAAAGCCCGCTCGTGCATTTGACCGGACAGAGATCCGCGGAGAGCGGGGCTG	14991
OY	3121	GACGCTTACGCGCGGTGCGCGTACGCTCGAACCTTCTTGAGGCGCCGATTCGCTTGT	3180

Db 14992 GACCTAAGCGCGCTGCGCGTAGCGTCCGAACGTTCTTGAAGCCCCGATTCGGTTGTT 15051  
Qy 3181 GAGTTTGGTCGATTCCTGAGCTGCTTGAGACAGCGTGAAGCCGAGCGGCGGACCTTCC 3240  
Db 15052 GAGTTTGGTCGATTCCTGAGCTGCTTGAGACAGCGTGAAGCCGAGCGGCGGACCTTCC 15111  
Qy 3241 AAATTCGTTATCCATCGGCGGCGGAGCAGTACCCGGTGAACAACTACCGGTATGCAAA 3300  
Db 15112 AAATTCGTTATCCATCGGCGGCGGAGCAGTACCCGGTGAACAACTACCGGTATGCAAA 15171  
Qy 3301 TCCGCGCCCATTCGAGGCGGTGAGACGAGGGCTTCTATTATTAACACCCGTTGAGACCGT 3360  
Db 15172 TCCGCGCCCATTCGAGGCGGTGAGACGAGGGCTTCTATTATTAACACCCGTTGAGACCGT 15231  
Qy 3361 TTGCTGAAGCTCTCCGATCAGCGGATTCGAGCGCGGAGCGACGTTCCGCAAACTTCGAC 3420  
Db 15232 TTGCTGAAGCTCTCCGATCAGCGGATTCGAGCGCGGAGCGACGTTCCGCAAACTTCGAC 15291  
Qy 3421 GTGTTGATGAAGCGCGCTTCAACCTCTGTTCTGTTGAGGAGAGATCGACCGCATCGAGTCG 3480  
Db 15292 GTGTTGATGAAGCGCGCTTCAACCTCTGTTCTGTTGAGGAGAGATCGACCGCATCGAGTCG 15351  
Qy 3481 CTGATGATGATGCTGCTCGCGAAGATTTTGGCTGCTGAGAGCGGATATATGCGGAGCTC 3540  
Db 15352 CTGATGATGATGCTGCTCGCGAAGATTTTGGCTGCTGAGAGCGGATATATGCGGAGCTC 15411  
Qy 3541 CTGATGAGACGAGCGCTTCTCTGCAACATCGGCGCTGCTGCGGTGGGCAATTCATTTT 3600  
Db 15412 CTGATGAGACGAGCGCTTCTCTGCAACATCGGCGCTGCTGCGGTGGGCAATTCATTTT 15471  
Qy 3601 GAAACGTTGCGCGCGGTTCTGCACTGCGACATTCGACGTTTACGTCGACGCGCATGCTG 3660  
Db 15472 GAAACGTTGCGCGCGGTTCTGCACTGCGACATTCGACGTTTACGTCGACGCGCATGCTG 15531  
Qy 3661 GCGCGGCGGCTGAGACCCCGCGGAGTTCCAGGTCTGTAGCTGCTGCTGCTGCTGCTGCTG 3720  
Db 15532 GCGCGGCGGCTGAGACCCCGCGGAGTTCCAGGTCTGTAGCTGCTGCTGCTGCTGCTGCTG 15591  
Qy 3721 AGGCGCGGCGGAGACCCCGCGGAGTTCCAGGTCTGTAGCTGCTGCTGCTGCTGCTGCTG 3780  
Db 15592 AGGCGCGGCGGAGACCCCGCGGAGTTCCAGGTCTGTAGCTGCTGCTGCTGCTGCTGCTG 15651  
Qy 3781 GCGGACTTCTTGAGAGCAAACTACCCGAGTACATGATGCTTACAGTCTTCTGAGAGCTC 3840  
Db 15652 GCGGACTTCTTGAGAGCAAACTACCCGAGTACATGATGCTTACAGTCTTCTGAGAGCTC 15711  
Qy 3841 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3900  
Db 15712 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 15771  
Qy 3901 GATACCTGCTGCGCGCGGATTCGCGGAGCAACGCGGAGCGCTTGGAGGAGATC 3960  
Db 15772 GATACCTGCTGCGCGCGGATTCGCGGAGCAACGCGGAGCGCTTGGAGGAGATC 15831  
Qy 3961 CTGCTGCGGCTGCTGCTGCGGAGTCTGCGGCTGAGAGTCTGCGGCTCAGAGAGCTTC 4020  
Db 15832 CTGCTGCGGCTGCTGCTGCGGAGTCTGCGGCTGAGAGTCTGCGGCTCAGAGAGCTTC 15891  
Qy 4021 GTTCATCTTGGTGGCAATTCATTCATTCGTCGATGAGAGGCTTGGTGAAGAGG 4080  
Db 15892 GTTCATCTTGGTGGCAATTCATTCATTCGTCGATGAGAGGCTTGGTGAAGAGG 15951  
Qy 4081 CTGATAGGAGATGCGCATCAGAGTGTTCAGTACCCGAACTCGGCTCGTGGCG 4140  
Db 15952 CTGATAGGAGATGCGCATCAGAGTGTTCAGTACCCGAACTCGGCTCGTGGCG 16011  
Qy 4141 TCCGTTTGGCGCGAGCTCGAGAGTGAATCAGCGCGCGAATCAGAGACCGAGT 4200  
Db 16012 TCCGTTTGGCGCGAGCTCGAGAGTGAATCAGCGCGCGAATCAGAGACCGAGT 16071  
Qy 4201 GAGGTTGCGCGCGAGAGGAGAGAGAGTGA 4233  
Db 16072 GAGGTTGCGCGCGAGAGGAGAGAGTGA 16104

RESULT 4  
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6355458.  
ACCESSION AR199559  
VERSION AR199559.1 GI:20249633  
KEYWORDS  
SOURCE  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)  
AUTHORS Schupp,T., Liqon,J, Madison., Molnar,I., Zirkle,R., Cyr,D, Dawn. and  
Gorlach,J.  
TITLE Genes for the biosynthesis of epoethlones  
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;  
FEATURES  
source 1..68750  
/organism="unknown"  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN  
Query Match 100.0%; Score 4233; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGACGATCATTCAGCTTCTGAGAGCTGAGACGAGGATATCAAGCTGGCGCGAT 60  
Db 11872 ATGACGATCATTCAGCTTCTGAGAGCTGAGACGAGGATATCAAGCTGGCGCGAT 11931  
Qy 61 GGGAGCGCCTCCAGATACAGGCGCCCAAGAACGCCCTGAACCCGAACCTGCTGCTGA 120  
Db 11932 GGGAGCGCCTCCAGATACAGGCGCCCAAGAACGCCCTGAACCCGAACCTGCTGCTGA 11991  
Qy 121 ATCTCCAGACAAAGACGATCTTACGATGCTCTGTCAGAGACTTCCCGCAGATTC 180  
Db 11992 ATCTCCAGACAAAGACGATCTTACGATGCTCTGTCAGAGACTTCCCGCAGATTC 12051  
Qy 181 ATCTGCCCCCGCCAGCGGAGCGGACGCTCCGTTTCTCTCAAGACATCAAGAAATCC 240  
Db 12052 ATCTGCCCCCGCCAGCGGAGCGGACGCTCCGTTTCTCTCAAGACATCAAGAAATCC 12111  
Qy 241 TACTGCTGGGCGGAGACGAGCTTTACGGTCCCGAGCGGATTCAGCCTATCGCAA 300  
Db 12112 TACTGCTGGGCGGAGACGAGCTTTACGGTCCCGAGCGGATTCAGCCTATCGCAA 12171  
Qy 301 TACGACTGTACGATCTTCGAGCTGCGAGGCTGAGCGCGCTTTCGAAAGTCTGCG 360  
Db 12172 TACGACTGTACGATCTTCGAGCTGCGAGGCTGAGCGCGCTTTCGAAAGTCTGCG 12231  
Qy 361 CCGACACGATGCTTTCGGGCGCCACAGCTGCGCCGACATGATGAGAGTGAACCTTAA 420  
Db 12232 CCGACACGATGCTTTCGGGCGCCACAGCTGCGCCGACATGATGAGAGTGAACCTTAA 12291  
Qy 421 GTGACGCGGACATCGAGATCATGATCTGCGGCGCTGCGACCGGAGCAGCAGGAAAGG 480  
Db 12292 GTGACGCGGACATCGAGATCATGATCTGCGGCGCTGCGACCGGAGCAGCAGGAAAGG 12351  
Qy 481 AGGCTGCTGTTGCGAGATGCGATGTCGACCGCATCTATYACACCGAGCGCCCTCG 540  
Db 12352 AGGCTGCTGTTGCGAGATGCGATGTCGACCGCATCTATYACACCGAGCGCCCTCG 12411  
Qy 541 CTCTATACGCTGCTGCGCTTTCGGCTGAGAGGCGGAAACCGCTCTGCTGCTGCTGCTG 600  
Db 12412 CTCTATACGCTGCTGCGCTTTCGGCTGAGAGGCGGAAACCGCTCTGCTGCTGCTGCTG 12471  
Qy 601 GATCTCATTAAGTTGAGCTTGAAGAGCTGCTCATCTTCAAGACTGCTGAGCTTC 660  
Db 12472 GATCTCATTAAGTTGAGCTTGAAGAGCTGCTCATCTTCAAGACTGCTGAGCTTC 12531  
Qy 661 TACGAAATCCCGAGACTCTCTCTGCTCTGAGGCTCTGTAACCGGATTAATGATCTC 720

Db 12532 TAGAAGATCCGAGACCTCTCTCCCTGTCCTGAGCTCTCGATCCGCGATTATGATCTC 12591  
QY 721 GCGCTGAGCTCTGCAAGAAAGTCTGAGCGCATCAACGATCGATGATTAATGGAACGG 780  
Db 12592 GCGCTGAGCTCTGCAAGAAAGTCTGAGCGCATCAACGATCGATGATTAATGGAACGG 12651  
QY 781 CCGATCGCGAGCTCCCACTCCCGCGAGAGCTTCCGATGAGAGCCGATCCATCTACCTG 840  
Db 12652 CCGATCGCGAGCTCCCACTCCCGCGAGAGCTTCCGATGAGAGCCGATCCATCTACCTG 12711  
QY 841 AAGGAGATCCGCTTCGCGACAAGAGCAATGGCTCCGTCGGACTCTGAGGTCGATTG 900  
Db 12712 AAGGAGATCCGCTTCGCGACAAGAGCAATGGCTCCGTCGGACTCTGAGGTCGATTG 12771  
QY 901 AAGCGGCTGTGCGGAGAGCGCGGCTGACCCCGACGCGCGTCATCTGCTGATTTTCC 960  
Db 12772 AAGCGGCTGTGCGGAGAGCGCGGCTGACCCCGACGCGCGTCATCTGCTGATTTTCC 12831  
QY 961 GAGTGATCGGCGCTGAGAGCGCGAGCCCGGTTTACGCTCAACATAAGCTCTTCAAC 1020  
Db 12832 GAGTGATCGGCGCTGAGAGCGCGAGCCCGGTTTACGCTCAACATAAGCTCTTCAAC 12891  
QY 1021 CCGCTCCCGCTCCATCCGCGCGTGAAGATATCAAGGAGACTTCAAGTGTGCTC 1080  
Db 12892 CCGCTCCCGCTCCATCCGCGCGTGAAGATATCAAGGAGACTTCAAGTGTGCTC 12951  
QY 1081 CTGGACATGACACCACTGCGCAAGAGCTTGAACAGCGCGCTAAGCGTATTCAGAG 1140  
Db 12952 CTGGACATGACACCACTGCGCAAGAGCTTGAACAGCGCGCTAAGCGTATTCAGAG 13011  
QY 1141 CAGCTGTGGAGAGCGATGATCACTGCAAGCTAAGGCTATCGAGTCCAGCGAGAGGCC 1200  
Db 13012 CAGCTGTGGAGAGCGATGATCACTGCAAGCTAAGGCTATCGAGTCCAGCGAGAGGCC 13071  
QY 1201 GCCCGGCTCTGCGGAGATCCAAAGAGCGGATTTCCCGTGTGCTCAAGCGGCTT 1260  
Db 13072 GCCCGGCTCTGCGGAGATCCAAAGAGCGGATTTCCCGTGTGCTCAAGCGGCGCTT 13131  
QY 1261 AACCAAGAGTCTGTGTGTCACTCTGTTCAGAGGCTGGAACCTCGGTGTAACACAC 1320  
Db 13132 AACCAAGAGTCTGTGTGTCACTCTGTTCAGAGGCTGGAACCTCGGTGTAACACAC 13191  
QY 1321 AGCGACATCTAGAGTGTGCTGATCACTCACTGAGAGCAAGATGGAGACCTGCTC 1380  
Db 13192 AGCGACATCTAGAGTGTGCTGATCACTCACTGAGAGCAAGATGGAGACCTGCTC 13251  
QY 1381 CTCGCTGGGACATCTGTGACGAGAGTGTCCCGCCGACCTTCTGAGACGATCTCGAA 1440  
Db 13252 CTCGCTGGGACATCTGTGACGAGAGTGTCCCGCCGACCTTCTGAGACGATCTCGAA 13311  
QY 1441 GCGTACGTGTTTTTCTCCGCGGCTCACTGAGGAACATGGGGTGAACAGTGGCTGT 1500  
Db 13312 GCGTACGTGTTTTTCTCCGCGGCTCACTGAGGAACATGGGGTGAACAGTGGCTGT 13371  
QY 1501 TCGCTTCGCGCTCCAGCTAGAAAGCGCGGCGAGGCAAAAGCGACAAAGCGCTGCTG 1560  
Db 13372 TCGCTTCGCGCTCCAGCTAGAAAGCGCGGCGAGGCAAAAGCGACAAAGCGCTGCTG 13431  
QY 1561 AGCGACATACGCTGACGCGCTGTTCGCGGCGCGGCTGAGAGCTGCCATGCACTC 1620  
Db 13432 AGCGACATACGCTGACGCGCTGTTCGCGGCGCGGCTGAGAGCTGCCATGCACTC 13491  
QY 1621 GCGCTGTGTGCGCGCAAGACGCTCAAGTGAAGAGCTTTCGCGCGTTCGCGCGA 1680  
Db 13492 GCGCTGTGTGCGCGCAAGACGCTCAAGTGAAGAGCTTTCGCGCGTTCGCGCGA 13551  
QY 1681 CTGGGCGCGCGCTGCGAGAGCGGGGCAAGCGCGCAACATTTGTCGCGGTGATG 1740  
Db 13552 CTGGGCGCGCGCTGCGAGAGCGGGGCAAGCGCGCAACATTTGTCGCGGTGATG 13611  
QY 1741 GAGAAAGCTGAGAGAGTGTTCGCGGTTCGCGGTGCTCGAGTCAGGCGCGGCTAC 1800  
Db 13612 GAGAAAGCTGAGAGAGTGTTCGCGGTTCGCGGTGCTCGAGTCAGGCGCGGCTAC 13671

QY 1801 GTGCCGATCGATGCCGACCTTACCGGCGGAGCGTATCCACTACCTCTCGATCATGTGAG 1860  
Db 13672 GTGCCGATCGATGCCGACCTTACCGGCGGAGCGTATCCACTACCTCTCGATCATGTGAG 13721  
QY 1861 GTAAAGCTGTGCTGACGAGCATGGCTGTGATGGAACATGTCATGAGCCGCGGGATC 1920  
Db 13732 GTAAAGCTGTGCTGACGAGCATGGCTGTGATGGAACATGTCATGAGCCGCGGGATC 13791  
QY 1921 CAGCGGCTCTGTGAGCGAGCGCGCTCGAAGCGACGCGGACCACTCTCGATGATG 1980  
Db 13792 CAGCGGCTCTGTGAGCGAGCGCGCTCGAAGCGACGCGGACCACTCTCGATGATG 13851  
QY 1981 CCCATTGACACCTTCCGATCTCGCGTATGATCTACACTCGGATTCACAGGCTT 2040  
Db 13852 CCCATTGACACCTTCCGATCTCGCGTATGATCTACACTCGGATTCACAGGCTT 13911  
QY 2041 CCCAAGGGGATGATGATCATCGAGGATGCGGCTCAACACATCTCTGATCAACAG 2100  
Db 13912 CCCAAGGGGATGATGATCATCGAGGATGCGGCTCAACACATCTCTGATCAACAG 13971  
QY 2101 CGCTTGAATAGGCGCGGAGACAGGCTGCGCTCTCTCTGCTGATGATCTC 2160  
Db 13972 CGCTTGAATAGGCGCGGAGACAGGCTGCGCTCTCTCTGCTGATGATCTC 14031  
QY 2161 TCGCTCTATGATGTGTGCGGATCTCTGCGCGCGGCTGATGATGTCGCGGACGG 2220  
Db 14032 TCGCTCTATGATGTGTGCGGATCTCTGCGCGCGGCTGATGATGTCGCGGACGG 14091  
QY 2221 TCCAAGCTGCGGATCCGCGCATGAGGCAAGTGTGATCGAAAGAGAGTGAAGG 2280  
Db 14092 TCCAAGCTGCGGATCCGCGCATGAGGCAAGTGTGATCGAAAGAGAGTGAAGG 14151  
QY 2281 TGAACCTGCTGCGGCGCTGATGCGATGCTGTGAGCACTTTGAGGCTGCCCGAT 2340  
Db 14152 TGAACCTGCTGCGGCGCTGATGCGATGCTGTGAGCACTTTGAGGCTGCCCGAT 14211  
QY 2341 TCGCTCTGATGCTCTGCGGCTTCTGCTGAGCGCGGCTGATGATGTCGCGGAGG 2400  
Db 14212 TCGCTCTGATGCTCTGCGGCTTCTGCTGAGCGCGGCTGATGATGTCGCGGAGG 14271  
QY 2401 CCGCGAGCTTCAGGCGATCAGGCGCGGCGGCTGTGATGATGAGCTGCGCGGCGCAC 2460  
Db 14272 CCGCGAGCTTCAGGCGATCAGGCGCGGCGGCTGTGATGATGAGCTGCGCGGCGCAC 14331  
QY 2461 GAAAGCTGATGTGCTCACTGCGGATCCCGTGAAGAACTGATCTGTGCGCGAGC 2520  
Db 14332 GAAAGCTGATGTGCTCACTGCGGATCCCGTGAAGAACTGATCTGTGCGCGAGC 14391  
QY 2521 ATCCCTACGCGCGCTCCGCTGCGCAACAGAGCTTCAAGCTGATGAGGCGCTCGAA 2580  
Db 14392 ATCCCTACGCGCGCTCCGCTGCGCAACAGAGCTTCAAGCTGATGAGGCGCTCGAA 14451  
QY 2581 CCGCGCGGCTGTGAGTTCCGCGGCAACTTACATTTGCGGCGTGTGCACTGTGAGC 2640  
Db 14452 CCGCGCGGCTGTGAGTTCCGCGGCAACTTACATTTGCGGCGTGTGCACTGTGAGC 14511  
QY 2641 TACTGCGGATGAGAGAGAGCGCGCAAGACTTCTGTGCACTCCGAGACCGGAG 2700  
Db 14512 TACTGCGGATGAGAGAGAGCGCGCAAGACTTCTGTGCACTCCGAGACCGGAG 14571  
QY 2701 CCGCTCTACAGACCGGCGATGTCGCGCGCTACCTGCGCGAGGAAACATGAGTTCA 2760  
Db 14572 CCGCTCTACAGACCGGCGATGTCGCGCGCTACCTGCGCGAGGAAACATGAGTTCA 14631  
QY 2761 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATACCGGCTGAGCTCGGAGAAATCGAG 2820  
Db 14632 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATACCGGCTGAGCTCGGAGAAATCGAG 14691  
QY 2821 GAAAGCTCAAGTGCATTCGAAGCTGACGCGAGCTGATGTCGCTCGGAAACGAC 2880  
Db 14692 GAAAGCTCAAGTGCATTCGAAGCTGACGCGAGCTGATGTCGCTCGGAAACGAC 14751



2881 GCGGCAACAAAGCTCTTCTAGCTATGTGTCTCCGAGGCGACACGAGACGCGCTCC 2940  
Db 14752 GCGGCAACAAAGCTCTTCTAGCTATGTGTCTCCGAGGCGACACGAGACGCGCTCC 14811  
Qy 2941 GAGCAGGACGAGGCTCTTAAGCCGAGGAGATTCGACGCGAGGACACGCGCCCGAAGC 3000  
Db 14812 GAGCAGGACGAGGCTCTTAAGCCGAGGAGATTCGACGCGAGGACACGCGCCCGAAGC 14871  
Qy 3001 GACGCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3060  
Db 14872 GACGCTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 14931  
Qy 3061 GACCTGAGCAGAAACCCGCTGCTGATTCGACCGGAGCAGATCCGCGAGGCGGAGCCTG 3120  
Db 14932 GACCTGAGCAGAAACCCGCTGCTGATTCGACCGGAGCAGATCCGCGAGGCGGAGCCTG 14991  
Qy 3121 GACCTGAGCAGAAACCCGCTGCTGATTCGACCGGAGCAGATCCGCGAGGCGGAGCCTG 3180  
Db 14992 GACCTGAGCAGAAACCCGCTGCTGATTCGACCGGAGCAGATCCGCGAGGCGGAGCCTG 15051  
Qy 3181 GAGTTGGTCAATTCCTGAGCTGCTTGAAGCAGCGAGCAGCGAGCAGCGAGCAGCGAGC 3240  
Db 15052 GAGTTGGTCAATTCCTGAGCTGCTTGAAGCAGCGAGCAGCGAGCAGCGAGCAGCGAGC 15111  
Qy 3241 AAATTCGCTTATTCATCGGCGGAGCAGCAGTACCGGATGCAAACTACCGGATGTCAAA 3300  
Db 15112 AAATTCGCTTATTCATCGGCGGAGCAGCAGTACCGGATGCAAACTACCGGATGTCAAA 15171  
Qy 3301 TCCGCGCCGATTCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3360  
Db 15172 TCCGCGCCGATTCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 15231  
Qy 3361 TTGCTGAAGCTCTCCGATCAACGAGATTCGAGCGAGGAGCAGCGAGGAGGAGGAGGAG 3420  
Db 15232 TTGCTGAAGCTCTCCGATCAACGAGATTCGAGCGAGGAGCAGCGAGGAGGAGGAGGAG 15291  
Qy 3421 GTGTTGATGAAGCGGCGCTTCAACTCTGTTGCTGAGGAGGAGGAGGAGGAGGAGGAG 3480  
Db 15292 GTGTTGATGAAGCGGCGCTTCAACTCTGTTGCTGAGGAGGAGGAGGAGGAGGAGGAG 15351  
Qy 3481 CTGTTGATGAAGCGGCGCTTCAACTCTGTTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 3540  
Db 15352 CTGTTGATGAAGCGGCGCTTCAACTCTGTTGCTGAGGAGGAGGAGGAGGAGGAGGAGG 15411  
Qy 3541 CTGATGAGCAGGCGCTTCCGCAACATTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 3600  
Db 15412 CTGATGAGCAGGCGCTTCCGCAACATTCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 15471  
Qy 3601 GAACAGGTTCCGCGCGGTTCTCGACCTGCGACATTCGAGCGGAGGAGGAGGAGGAGGAG 3660  
Db 15472 GAACAGGTTCCGCGCGGTTCTCGACCTGCGACATTCGAGCGGAGGAGGAGGAGGAGGAG 15531  
Qy 3661 GCGCGGCGGAGTGAACCCGCGGAGTTCAGAGTCTGTAAGCTCGTGCAGAGATTCCTACCG 3720  
Db 15532 GCGCGGCGGAGTGAACCCGCGGAGTTCAGAGTCTGTAAGCTCGTGCAGAGATTCCTACCG 15591  
Qy 3721 AGGCGCGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3780  
Db 15592 AGGCGCGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 15651  
Qy 3781 CGCGACTTCTTGAGAGCAAACTACCGAGTATCATGTGCTTCAAGCTCTTCTGAGGCTC 3840  
Db 15652 CGCGACTTCTTGAGAGCAAACTACCGAGTATCATGTGCTTCAAGCTCTTCTGAGGCTC 15711  
Qy 3841 GATGCTTGCGGCTGAGCTCAACGAGCAAGTGCATTCGTAAGCGCTTCGCGAGCGAGAG 3900  
Db 15712 GATGCTTGCGGCTGAGCTCAACGAGCAAGTGCATTCGTAAGCGCTTCGCGAGCGAGAG 15771  
Qy 3901 GATACCTTGCGGCGGCGGATTCGAGGAGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3960  
Db 15772 GATACCTTGCGGCGGCGGATTCGAGGAGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 15831  
Qy 3961 CTGCTGCGGCTGTAACGAGAGGTGCTCGGCGTGAAGGTGCTCGGCGCTCAGAGAGCTTC 4020

15832 CTCCTCGGCTGTAACGAGAGTCTCGGCTGAGAGGTGTCGAGCTTCACAGAGCTTC 15891  
Qy 4021 GTGATCTTGAGAGATTCGATTCATCATGCTTCGATGAGAGGCTGTCAGAGAGG 4080  
Db 15892 GTGATCTTGAGAGATTCGATTCATCATGCTTCGATGAGAGGCTGTCAGAGAGG 15951  
Qy 4081 CTGATGAGAGATTCGATTCATCATGCTTCGATGAGAGGCTGTCAGAGAGG 4140  
Db 15952 CTGATGAGAGATTCGATTCATCATGCTTCGATGAGAGGCTGTCAGAGAGG 16011  
Qy 4141 TCCGTTTGCGGCGAGATTCGAGAGTTCATGATCAGCGGCGGAGGAGGAGGAGGAGG 4200  
Db 16012 TCCGTTTGCGGCGAGATTCGAGAGTTCATGATCAGCGGCGGAGGAGGAGGAGGAGG 16071  
Qy 4201 GAGGTTGCGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4233  
Db 16072 GAGGTTGCGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 16104

RESULT 5  
AR199567  
LOCUS AR199567 68750 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 635459.  
ACCESSION AR199567  
VERSION AR199567.1 GI:20249641  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
BASE COUNT  
ORIGIN

Query Match 100.0%; Score 4233; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGACGATCATGCTTCTGAAGCTGAGACGAGGATTCAGAGCTGCGGAGATCC 60  
Db 11872 ATGACGATCATGCTTCTGAAGCTGAGACGAGGATTCAGAGCTGCGGAGATCC 11931  
Qy 61 GGGAGCGCTCCAGATTCAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120  
Db 11932 GGGAGCGCTCCAGATTCAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11991  
Qy 121 ATTCGAGCAGCAAAAGCAGATTCGAGAGTCTCGTCAAGAGCTCCCGCAGATCC 180  
Db 11992 ATTCGAGCAGCAAAAGCAGATTCGAGAGTCTCGTCAAGAGCTCCCGCAGATCC 12051  
Qy 181 ATTCGAGCAGCAAAAGCAGATTCGAGAGTCTCGTCAAGAGCTCCCGCAGATCC 240  
Db 12052 ATTCGAGCAGCAAAAGCAGATTCGAGAGTCTCGTCAAGAGCTCCCGCAGATCC 12111  
Qy 241 TACTGCTGAGGCGGAGCAGAGGCTTTCAGGCTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
Db 12112 TACTGCTGAGGCGGAGCAGAGGCTTTCAGGCTCCCGAGGAGGAGGAGGAGGAGGAGGAGG 12171  
Qy 301 TACGATGATGAGATTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAG 360  
Db 12172 TACGATGATGAGATTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAGTTCGAGAG 12231  
Qy 361 CGGACGAGATGCTTTCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420  
Db 12232 CGGACGAGATGCTTTCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12291



OY	421	GTGACGCCGACATCGAGATCATCGATCTCGGGGCTGACCGGAGCACAGGGAAAGG	480
Db	12292	GTGACGCCGACATCGAGATCATCGATCTCGGGGCTGACCGGAGCACAGGGAAAGG	12351
OY	481	AGGCTCGTGTGCTTGCGAGATGCGATTCGCGACCGCATCTATGACACGAGCGCCCTCCG	540
Db	12352	AGGCTCGTGTGCTTGCGAGATGCGATTCGCGACCGCATCTATGACACGAGCGCCCTCCG	12411
OY	541	CTCTATACGTCGTGCGCGTTCCGGCTGGACGAGCGCAAAACCCTCTCGTCTCAGTATC	600
Db	12412	CTCTATACGTCGTGCGCGTTCCGGCTGGACGAGCGCAAAACCCTCTCGTCTCAGTATC	12471
OY	601	GATTCCTATTAACGTTGACCTTAGAGAGCTTCATCATCTTCAAGGACTGCGTCACTTC	660
Db	12472	GATTCCTATTAACGTTGACCTTAGAGAGCTTCATCATCTTCAAGGACTGCGTCACTTC	12531
OY	661	TACGAAGATCCCGAGACCTCTCTCCCTGCTGTGAGCTCTCGAACCGATTATGATATC	720
Db	12532	TACGAAGATCCCGAGACCTCTCTCCCTGCTGTGAGCTCTCGAACCGATTATGATATC	12591
OY	721	GCCTGTGAGTCTCGCAAGAAAGTCTGAGCGCATCAACGATCGATGATTACTGGAACCG	780
Db	12592	GCCTGTGAGTCTCGCAAGAAAGTCTGAGCGCATCAACGATCGATGATTACTGGAACCG	12651
OY	781	CGCATCCGCCAGTCTCCACCTTCGCGCGACGCTTCGATGAAGCCGATCATCTACCTCG	840
Db	12652	CGCATCCGCCAGTCTCCACCTTCGCGCGACGCTTCGATGAAGCCGATCATCTACCTCG	12711
OY	841	AAGAGATATCCGCTCCGGACACAGAGACATAGGCTGCGCGATCTCGGGGTGCAATTG	900
Db	12712	AAGAGATATCCGCTCCGGACACAGAGACATAGGCTGCGCGATCTCGGGGTGCAATTG	12771
OY	901	AAGCGCGTGTCCGGGAGCGCGGCTGACCCCGACGCGCTCATCTCGCTGATATTTTC	960
Db	12772	AAGCGCGTGTCCGGGAGCGCGGCTGACCCCGACGCGCTCATCTCGGCTGATATTTTC	12831
OY	961	GAGGTGATCCGGGCGCTTGAGAGCGGAGGCCCGGTTTACGTCAACATACGCTTTCAC	1020
Db	12832	GAGGTGATCCGGGCGCTTGAGAGCGGAGGCCCGGTTTACGTCAACATACGCTTTCAC	12891
OY	1021	CGGCTCCCGGTCATCCGCGCGTGAAGGATATCAACGGGGACTTCACGTGATGGTCCTC	1080
Db	12892	CGGCTCCCGGTCATCCGCGCGTGAAGGATATCAACGGGGACTTCACGTGATGGTCCTC	12951
OY	1081	CTGGACATTCGACACCACTCGCGACAAAGAGCTTCGAAACAGGCGCTTAAGGTATTC	1140
Db	12952	CTGGACATTCGACACCACTCGCGACAAAGAGCTTCGAAACAGGCGCTTAAGGTATTC	13011
OY	1141	CAGCTGTGGGAGCGATGATCATCTGCGAGCTAAGCGGTATCGAGGTCCAGCGAGGCTC	1200
Db	13012	CAGCTGTGGGAGCGATGATCATCTGCGAGCTAAGCGGTATCGAGGTCCAGCGAGGCTC	13071
OY	1201	GCCCGGGGCTCGGGGGATTCACAAGAGAGGCAATTTGCTCCCGTGGTGTCAAGAGCGGCTT	1260
Db	13072	GCCCGGGGCTCGGGGGATTCACAAGAGAGGCAATTTGCTCCCGTGGTGTCAAGAGCGGCTT	13131
OY	1261	AACCAAGAGTCTGTGATCACCTCGTTGACAGAGGCTCGGAATCCCGTGTATCACACAGC	1320
Db	13132	AACCAAGAGTCTGTGATCACCTCGTTGACAGAGGCTCGGAATCCCGTGTATCACACAGC	13191
OY	1321	ACGCAAGACTCTCAAGCTGCTGCTGAGATCATCACTACCTTACGAGCAGATGGGACCTCGTC	1380
Db	13192	ACGCAAGACTCTCAAGCTGCTGCTGAGATCATCACTACCTTACGAGCAGATGGGACCTCGTC	13251
OY	1381	CTGCGCGGGGACATCGTGTGACGAGAGTGTCCCGCCGACCTTCTGGAACGATAGCTCGAA	1440
Db	13252	CTGCGCGGGGACATCGTGTGACGAGAGTGTCCCGCCGACCTTCTGGAACGATAGCTCGAA	13311
OY	1441	GCGTACGTGCTTTTCTCCGGCGGCTCATCTGAGAAACATAGGGGTAAACAGGTGCGCTGT	1500
Db	13312	GCGTACGTGCTTTTCTCCGGCGGCTCATCTGAGAAACATAGGGGTAAACAGGTGCGCTGT	13371
OY	1501	TGCGTTCGCGCTGCGCAACTGAAAGCGCGGGCAGACGCAACCGCGTCTTG	1560

Db	13372	TCGCTTCCGCTGCTGCCAGCTAGAAAGCGCGGAGCGCAACCGCAACCGCGCTGCTG	13433
Qy	1561	AGCGAGCATACGCTGCACGGCGCTTTCGCGCGCGGGGTGAGCAGCTGCCATGCACTC	1620
Db	13432	AGCGAGCATACGCTGCACGGCGCTGTTTCGCGCGCGGGGTGAGCAGCTGCCATGCACTC	13491
Qy	1621	GCCGTGCTGTCGCGCGCGCAAGACGCTTCACGTACGAAGAGCTTTCGCGCGCTTCGCGCGCA	1680
Db	13492	GCCGTGCTGTCGCGCGCGCAAGAGCTTCACGTACGAAGAGCTTTCGCGCGCTTCGCGCGCA	13551
Qy	1681	CTTGGCGCGGGCGCGCGAGCAGGGGGACCGCCGCAACATTTGGTCGCGGTGATG	1740
Db	13552	CTTGGCGCGGGCGCGCGAGCAGGGGGACCGCCGCAACATTTGGTCGCGGTGATG	13611
Qy	1741	GAGAAAGCTGGGAGCAGTGTGTGCGGTTCTGCGGTGCTGAGTCAGAGCGCGCTAC	1800
Db	13612	GAGAAAGCTGGGAGCAGTGTGTGCGGTTCTGCGGTGCTGAGTCAGAGCGCGCTAC	13672
Qy	1801	GTGCGCATCGATGCGCCTTACCGCGGAGCGTATCATCTCTCGATCATGTGAG	1860
Db	13672	GTGCGCATCGATGCGCCTTACCGCGGAGCGTATCATCTCTCGATCATGTGAG	13733
Qy	1861	GTAAGCTCGTGCAGACGCGCATGGCTGAGAGCAACCTGTCATAGGCGCGCGGGATC	1920
Db	13732	GTAAGCTCGTGCAGACGCGCATGGCTGAGAGCAACCTGTCATAGGCGCGCGGGATC	13791
Qy	1921	CAGCGGCTGCTCGTGACGAGCGCGGCTGAGAGCGACGCGCACCGCCTCGATGATG	1980
Db	13792	CAGCGGCTGCTCGTGACGAGCGCGGCTGAGAGCGACGCGCACCGCCTCGATGATG	13851
Qy	1981	CCCATTCAGACACCTTCGATCTTCGCGTATGATCATCACTCGGGATTCACAGGTTG	2040
Db	13852	CCCATTCAGACACCTTCGATCTTCGCGTATGATCATCACTCGGGATTCACAGGTTG	13911
Qy	2041	CCCAAGGGGGTGATGATTCGATCATCGGGGGCGGTACAACCATCTGGAACATCAAGAG	2100
Db	13912	CCCAAGGGGGTGATGATTCGATCATCGGGGGCGGTACAACCATCTGGAACATCAAGAG	13971
Qy	2101	CGCTTCGAAATAGGGGCCCGGAGACAGAGGTGCTGGCGCTCTCGCTGAGCTTCGATCTC	2160
Db	13972	CGCTTCGAAATAGGGGCCCGGAGACAGAGGTGCTGGCGCTCTCGCTGAGCTTCGATCTC	14031
Qy	2161	TCGGTCTATGATGTGTTGCGGAATCTGCGCGCGGCGGTACGATCGTGTGCCGACGCG	2220
Db	14032	TCGGTCTATGATGTGTTGCGGAATCTGCGCGCGGCGGTACGATCGTGTGCCGACGCG	14091
Qy	2221	TTCAGAGTCGCGCATTCGCGCGCATTTGGCAGAGTTTATCGAACGAGAGAGGTGACGGTG	2280
Db	14092	TTCAGAGTCGCGCATTCGCGCGCATTTGGCAGAGTTTATCGAACGAGAGAGGTGACGGTG	14151
Qy	2281	TGGAACCTCGGTGCGCGCGCTGAGACGGATGCTGTCGAGCATTTTGAAGGTCGCCCGAT	2340
Db	14152	TGGAACCTCGGTGCGCGCGCTGAGACGGATGCTGTCGAGCATTTTGAAGGTCGCCCGAT	14211
Qy	2341	TCGCTGCTAGGTCTCTGCGGCTTTTCGCTGCTGAGCGCGCATGATCCCGGTGGGCTTG	2400
Db	14212	TCGCTGCTAGGTCTCTGCGGCTTTTCGCTGCTGAGCGCGCATGATCCCGGTGGGCTTG	14271
Qy	2401	CTGCGGAGAGCTCCAGGCGCATCAAGGCCCGCGGTGTGCGTATCAAGCCTGGGGCGGGCACAC	2460
Db	14272	CTGCGGAGAGCTCCAGGCGCATCAAGGCCCGCGGTGTGCGTATCAAGCCTGGGGCGGGCACAC	14331
Qy	2461	GAAGCGTCGATCTGTGTCATCGGGTACCCCGTGAGAGACGTCACTATCTGTGGCGAGC	2520
Db	14332	GAAGCGTCGATCTGTGTCATCGGGTACCCCGTGAGAGACGTCACTATCTGTGGCGAGC	14391
Qy	2521	ATCCCTTACGAGCGGTCCGCTGCGCAACCAAGCGTTTACAGTGTGATGAGGCGCTCGAA	2580
Db	14392	ATCCCTTACGAGCGGTCCGCTGCGCAACCAAGCGTTTACAGTGTGATGAGGCGCTCGAA	14451
Qy	2581	CCGCGCCCGGTCTGGGTTCCGGGGCAACTTACATTGGCGGGGTCCGGCTGCGACCTGGGGC	2640

Db 14452 CCGGCCCCGTCCTGGGTTCCGGGCACTCTAATTGCGGGGTCGGCTGGCACTGGGC 14511  
 Qy 2641 TACTGGCGCGATGAGAGAGAGCGCGCAAGCTTCTCTGTCACCCCGAGACCGGGAG 2700  
 Db 14512 TACTGGCGCGATGAGAGAGAGCGCGCAAGCTTCTCTGTCACCCCGAGACCGGGAG 14571  
 Qy 2701 CGCCTTACAAAGACCGCGATCTGGGCGCTAAGCTTCCCGATGAGAAACATGAGTTCA 2760  
 Db 14572 CGCCTTACAAAGACCGCGATCTGGGCGCTAAGCTTCCCGATGAGAAACATGAGTTCA 14631  
 Qy 2761 GGGGCTGAGCAACCAATCAAGCTTGGCGGATACCGCGTTGAGCTGGGGAAATCGAG 2820  
 Db 14632 GGGGCTGAGCAACCAATCAAGCTTGGCGGATACCGCGTTGAGCTGGGGAAATCGAG 14691  
 Qy 2821 GAAAGCTCAAGTGCATCCGAACTGACGCGAGCGGATGATTTGTCGCCCTGGGAGAC 2880  
 Db 14692 GAAAGCTCAAGTGCATCCGAACTGACGCGAGCGGATGATTTGTCGCCCTGGGAGAC 14751  
 Qy 2881 GCGGCAACAGCTCTTCTAGCCTATGTCCTCCGAGGGGCAACGAGACGCGCTCC 2940  
 Db 14752 GCGGCAACAGCTCTTCTAGCCTATGTCCTCCGAGGGGCAACGAGACGCGCTCC 14811  
 Qy 2941 GAGCAGACGCGAGCTTCAAGCCGAGCGGATCCAGCGGAGCAACGCGCGGAGACG 3000  
 Db 14812 GAGCAGACGCGAGCTTCAAGCCGAGCGGATCCAGCGGAGCAACGCGCGGAGACG 14871  
 Qy 3001 GACGCTTGAAGCGAGCGGAGGGTGAAGTTCAAGCTCGCTCGACACGACTCCGAGG 3060  
 Db 14872 GACGCTTGAAGCGAGCGGAGGGTGAAGTTCAAGCTCGCTCGACACGACTCCGAGG 14931  
 Qy 3061 GACCTGAGCGGAAAGCCCGTCGTGATGACCGGGAGATCCGCGGAGCGCGGGCTG 3120  
 Db 14932 GACCTGAGCGGAAAGCCCGTCGTGATGACCGGGAGATCCGCGGAGCGCGGGCTG 14991  
 Qy 3121 GACGCTTGAAGCGCGCTCGCGTGAAGCTCCGTAAGCGCGGATTCCTGTTGTT 3180  
 Db 14992 GACGCTTGAAGCGCGCTCGCGTGAAGCTCCGTAAGCGCGGATTCCTGTTGTT 15051  
 Qy 3181 GAGTTTGGTGCATTCCTGAGCTGTTGAGACGCGTGAAGCGCGGAGCGCGGCTTCC 3240  
 Db 15052 GAGTTTGGTGCATTCCTGAGCTGTTGAGACGCGTGAAGCGCGGAGCGCGGCTTCC 15111  
 Qy 3241 AAATTCGCTTATCCATCGCGGCGGAGCAAGTACCCGCTGCAAACTTACGATGTCAA 3300  
 Db 15112 AAATTCGCTTATCCATCGCGGCGGAGCAAGTACCCGCTGCAAACTTACGATGTCAA 15171  
 Qy 3301 TCCGCGCCCATCGAGGCGTGAAGAGGCTTCTATTATACACCGCTTGAAGACCTG 3360  
 Db 15172 TCCGCGCCCATCGAGGCGTGAAGAGGCTTCTATTATACACCGCTTGAAGACCTG 15231  
 Qy 3361 TTGCTGAAGCTCTCCGATCAAGGAGTGAAGGCGGAGGCACTTCCGCAAACTTGCAC 3420  
 Db 15232 TTGCTGAAGCTCTCCGATCAAGGAGTGAAGGCGGAGGCACTTCCGCAAACTTGCAC 15291  
 Qy 3421 GTGTTGATGAAGCGCGCTTCAACTCTGTTCTGTTGAGGAGAGATGAGCCATCGAGTCG 3480  
 Db 15292 GTGTTGATGAAGCGCGCTTCAACTCTGTTCTGTTGAGGAGAGATGAGCCATCGAGTCG 15351  
 Qy 3481 CTGATGATGATGCTGTCGCGGAGATTTTCTGCTGCTGAGGCGCGATATATGCGCAGCTC 3540  
 Db 15352 CTGATGATGATGCTGTCGCGGAGATTTTCTGCTGCTGAGGCGCGATATATGCGCAGCTC 15411  
 Qy 3541 CTGATGAGCAAGCGCGCTTCTCGCAACATCGCGGCTGTCGCGTGGGCAATTCATTTT 3600  
 Db 15412 CTGATGAGCAAGCGCGCTTCTCGCAACATCGCGGCTGTCGCGTGGGCAATTCATTTT 15471  
 Qy 3601 GAACAGTTTCGCGCGCTTCTCGACCTGCGACATTCGACGTTTACGTCACGCGATGCTG 3660  
 Db 15472 GAACAGTTTCGCGCGCTTCTCGACCTGCGACATTCGACGTTTACGTCACGCGATGCTG 15531  
 Qy 3661 GCGCGGCGGATGAGACCGCGGCACTTCCAGGCTTCTACGCTCGGTCGAGATTCCTACCG 3720  
 Db 15532 GCGCGGCGGATGAGACCGCGGCACTTCCAGGCTTCTACGCTCGGTCGAGATTCCTACCG 15591

Qy 3721 AGCGCGCAAGACGCGCGGCGCCCTTCCCGCGCGAGACAGACTTTCGCGATATGCTT 3780  
 Db 15592 AGCGCGCGCAAGACGCGCGGCGCCCTTCCCGCGCGAGACAGACTTTCGCGATATGCTT 15651  
 Qy 3781 CGGCACTTCTTGAAGACCAACTACCGAGATGATGTCCTTACAGTCTTGTGAGCTC 3840  
 Db 15652 CGGCACTTCTTGAAGACCAACTACCGAGATGATGTCCTTACAGTCTTGTGAGCTC 15711  
 Qy 3841 GATGCTTCCGCTGAGTCCCAAGGAGGATGATGATGAGGCTTGGCGGAGCGAG 3900  
 Db 15712 GATGCTTCCGCTGAGTCCCAAGGAGGATGATGATGAGGCTTGGCGGAGCGAG 15771  
 Qy 3901 GATACCTGCTGCGCGGCACTTCCGGGCAACAGCGCGCAACGCGCTTGGAGAGATC 3960  
 Db 15772 GATACCTGCTGCGCGGCACTTCCGGGCAACAGCGCGCAACGCGCTTGGAGAGATC 15831  
 Qy 3961 CTGCTCGGCTGTAAGGAGGCTCTCGGCTGAGAGGTCGAGCTTCACAGACTTC 4020  
 Db 15832 CTGCTCGGCTGTAAGGAGGCTCTCGGCTGAGAGGTCGAGCTTCACAGACTTC 15891  
 Qy 4021 GTGATCTTGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 4080  
 Db 15892 GTGATCTTGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 15951  
 Qy 4081 CTGATAGGAGATGCGCATACCGAGTGTTCAGTACCGGAACCTTCTGCTGCGG 4140  
 Db 15952 CTGATAGGAGATGCGCATACCGAGTGTTCAGTACCGGAACCTTCTGCTGCGG 16011  
 Qy 4141 TCCGCTTTCGCGGAGACTGAGAGATCTAGATGAGCGCGGCAATGACGAGCCGAGT 4200  
 Db 16012 TCCGCTTTCGCGGAGACTGAGAGATCTAGATGAGCGCGGCAATGACGAGCCGAGT 16071  
 Qy 4201 GAGGTTGCGCGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4233  
 Db 16072 GAGGTTGCGCGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16104

RESULT 6  
 AR201097  
 LOCUS 68750 bp DNA linear PAT 20-APR-2002  
 DEFINITION Sequence 1 from patent US 6358719.  
 ACCESSION AR201097  
 VERSION AR201097.1 GI:20251985  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Unknown.  
 Unclassified.  
 1 (bases 1 to 68750)  
 REFERENCE  
 Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and  
 Gortlach,J.  
 TITLE Genes for the biosynthesis of epoethilones  
 JOURNAL Patent: US 6358719-A 1 19-Mar-2002;  
 FEATURES  
 source  
 1..68750  
 /organism="unknown"  
 BASE COUNT 9596 a 22456 c 25539 g 11159 t  
 ORIGIN  
 Query Match 100.0%; Score 4233; DB 6; Length 68750;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ATGACATCATCATGCTTCTGAAGAGCTGAGACCAAGGATATCAAGCTGGCGCGAT 60  
 Db 11872 ATGACATCATCATGCTTCTGAAGAGCTGAGACCAAGGATATCAAGCTGGCGCGAT 11931  
 Qy 61 GGGAGAGCGCTCCAGATACAGGCGCCCAAGAACGCTGAAACCGGAACCTGCTGCA 120  
 Db 11932 GGGAGAGCGCTCCAGATACAGGCGCCCAAGAACGCTGAAACCGGAACCTGCTGCA 11991  
 Qy 121 ATCTCGAGCACAAAAGACGATCTGACGATGCTCTGTCAGAGACTTCCCGCAGAAATCC 180

Db 11992 ATCTCCGACGACAAAAGCAGATCTGACGATGCTCCGTCAAGACTCCCGCGAAGATCC 12051  
Qy 181 ATCTGTCGCCCCCGACCCGAGCGGACGCTCGTTTCTCTCAAGACATCCAGATATCC 240  
Db 12052 ATCTGTCGCCCCCGACCCGAGCGGACGCTCCGTTTCTCTCAAGACATCCAGATATCC 12111  
Qy 241 TACTGGCTGGGCGGACAGAGGGTTTACGATCCCGAGGGAGTCCAGCCTATCGCGAA 300  
Db 12112 TACTGGCTGGGCGGACAGAGGGTTTACGATCCCGAGGGAGTCCAGCCTATCGCGAA 12171  
Qy 301 TACGACTGACGATCTCGACGTCGCGAGGCTGAGCGCGCTTTTGAAAGTCTGTCCG 360  
Db 12172 TACGACTGACGATCTCGACGTCGCGAGGCTGAGCGCGCTTTTGAAAGTCTGTCCG 12231  
Qy 361 CGGACAGACATGCTTGGGGCCACAGCTGCGCGACATGATGAGGTGATCCAGCTTAA 420  
Db 12232 CGGACAGACATGCTTGGGGCCACAGCTGCGCGACATGATGAGGTGATCCAGCTTAA 12291  
Qy 421 GTCCAGCGCGACATCGAGATCATCGATCTGCGGGGCTGACCGGAGCAGCGGAGGCG 480  
Db 12292 GTCCAGCGCGACATCGAGATCATCGATCTGCGGGGCTGACCGGAGCAGCGGAGGCG 12351  
Qy 481 AGGCTGTGTCTGTGGAGATGCGATGTGCAACCGCATATGACACCGAGCGCCCTCCG 540  
Db 12352 AGGCTGTGTCTGTGGAGATGCGATGTGCAACCGCATATGACACCGAGCGCCCTCCG 12411  
Qy 541 CTCTATCAAGTGTGTGCGCTTGGGCTGAGACAGCGGCAAAACCGTCTCTGTCTCAATTC 600  
Db 12412 CTCTATCAAGTGTGTGCGCTTGGGCTGAGACAGCGGCAAAACCGTCTCTGTCTCAATTC 12471  
Qy 601 GATCTCATTAAGTTGACCTTGAAGCAGCTGTCCATCATCTTCAAGAGCTGGCTCACTTC 660  
Db 12472 GATCTCATTAAGTTGACCTTGAAGCAGCTGTCCATCATCTTCAAGAGCTGGCTCACTTC 12531  
Qy 661 TACGAAGATCCGAGACCTCTCTCCCTGTCTGAGACTCTGTGACCGCATATGATATC 720  
Db 12532 TACGAAGATCCGAGACCTCTCTCCCTGTCTGAGACTCTGTGACCGCATATGATATC 12591  
Qy 721 GGGCTGGAATCTCGCAAGAACTGTGAGGCGCATCAACGATCGATTAAGTGAAGCGG 780  
Db 12592 GGGCTGGAATCTCGCAAGAACTGTGAGGCGCATCAACGATCGATTAAGTGAAGCGG 12651  
Qy 781 CGCATGCGCGAGCTCCACCTCGCGCGAGCGCTTCGATGAAGGCGCATCATACCTTCG 840  
Db 12652 CGCATGCGCGAGCTCCACCTCGCGCGAGCGCTTCGATGAAGGCGCATCATACCTTCG 12711  
Qy 841 AAGAGATCCGCTTCCGACACAGAGCAATGCTGCGACTCTGAGGGTCGATTC 900  
Db 12712 AAGAGATCCGCTTCCGACACAGAGCAATGCTGCGACTCTGAGGGTCGATTC 12771  
Qy 901 AAGCGCGGTGTGGGGAGCGCGGGCTGACCGGCGGTGATCTGTGCTGATTTTC 960  
Db 12772 AAGCGCGGTGTGGGGAGCGCGGGCTGACCGGCGGTGATCTGTGCTGATTTTC 12831  
Qy 961 GAGGTGATCGGGGCTGAGACGCGAGCCCGGTTTACGCTCAACATTAAGCTTTCAAC 1020  
Db 12832 GAGGTGATCGGGGCTGAGACGCGAGCCCGGTTTACGCTCAACATTAAGCTTTCAAC 12891  
Qy 1021 CGGCTCCCGGTTCATCGCGCGTGAACGATATCACCGGGGACTTCAAGTGTCTTC 1080  
Db 12892 CGGCTCCCGGTTCATCGCGCGTGAACGATATCACCGGGGACTTCAAGTGTCTTC 12951  
Qy 1081 CTGGAATGACACCACTCGCGACAGAGCTTGAACAGCGGCTTAAGGTATTCAGAG 1140  
Db 12952 CTGGAATGACACCACTCGCGACAGAGCTTGAACAGCGGCTTAAGGTATTCAGAG 13011  
Qy 1141 CAGCTGTGGAAAGCATGTATCTGCAAGTAAAGGCTATCGAGGTCCAGCGAGGCG 1200  
Db 13012 CAGCTGTGGAAAGCATGTATCTGCAAGTAAAGGCTATCGAGGTCCAGCGAGGCG 13071  
Qy 1201 GCCCGGGTCTGTGGGATCCACAGAGCGCATTTGTCCTCGGTGTCTCAGAGCGGCTT 1260  
Db 13072 GCCCGGGTCTGTGGGATCCACAGAGCGCATTTGTCCTCGGTGTCTCAGAGCGGCTT 13131

Qy 1261 AACGAGAGTGTGTGTGACCTCGTTGAGAGAGGCTCGGAACCTCCGCTGACACGAGC 1320  
Db 13132 AACGAGAGTGTGTGTGACCTCGTTGAGAGAGGCTCGGAACCTCCGCTGACACGAGC 13191  
Qy 1321 ACGGAGCTCTCAGCTCTGTGATCATGACTCTACGAGCAGATGGGAGCTCTGTC 1380  
Db 13192 ACGGAGCTCTCAGCTCTGTGATCATGACTCTACGAGCAGATGGGAGCTCTGTC 13251  
Qy 1381 CTGCGGTGGACATGTGTGACGGAATGTTCCCGCCGACCTTTGTGAGACATGCTCGAA 1440  
Db 13252 CTGCGGTGGACATGTGTGACGGAATGTTCCCGCCGACCTTTGTGAGACATGCTCGAA 13311  
Qy 1441 GGTAGGTGTGTTTTCTCCGGGGCTCATCTGAGAACCATGGGGGAAACAGGTGGCTGT 1500  
Db 13312 GGTAGGTGTGTTTTCTCCGGGGCTCATCTGAGAACCATGGGGGAAACAGGTGGCTGT 13371  
Qy 1501 TCGCTTCCGCTGCCAGCTAGAAAGCGGGGAGCGGAAACGCGACCAACGCGCTGTG 1560  
Db 13372 TCGCTTCCGCTGCCAGCTAGAAAGCGGGGAGCGGAAACGCGACCAACGCGCTGTG 13431  
Qy 1561 ACGGAGCATACGCTGACAGGCTGTTCCGGGCGCGGCTGAGACAGCTGCCATGACCTC 1620  
Db 13432 ACGGAGCATACGCTGACAGGCTGTTCCGGGCGCGGCTGAGACAGCTGCCATGACCTC 13491  
Qy 1621 GCGGTGTGTGCGGCGCGGAGAGCGCTCAAGTACGAAAGCTTTGCGGCGCTTCCGGGCG 1680  
Db 13492 GCGGTGTGTGCGGCGCGGAGAGCGCTCAAGTACGAAAGCTTTGCGGCGCTTCCGGGCG 13551  
Qy 1681 CTGTGCGCGGCTCGGAGAGAGGAGGACGCGGAAACATTTGTCGCGGTGTGTATG 1740  
Db 13552 CTGTGCGCGGCTCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13611  
Qy 1741 GAGAAAGCTGGGAGAGAGGTTGTCCCGGTTCTCGGCTGCTGAGTCAAGCGCGGCTTAC 1800  
Db 13612 GAGAAAGCTGGGAGAGAGGTTGTCCCGGTTCTCGGCTGCTGAGTCAAGCGCGGCTTAC 13671  
Qy 1801 GTGCGGATCGAGCGGACCTTACCGGGGAGGCTATTCATACCTGCTGATGATGTTGAG 1860  
Db 13672 GTGCGGATCGAGCGGACCTTACCGGGGAGGCTATTCATACCTGCTGATGATGTTGAG 13731  
Qy 1861 GTAAAGCTGTGTCTGACGAGCAGCTGATGATGACCACTGTCAATGAGCGCGGAGATC 1920  
Db 13732 GTAAAGCTGTGTCTGACGAGCAGCTGATGATGACCACTGTCAATGAGCGCGGAGATC 13791  
Qy 1921 CAGCGGCTGTGTGAGAGAGGCGCGGCTTCAAGGCGAGCGGACAGCTTCCATGATG 1980  
Db 13792 CAGCGGCTGTGTGAGAGAGGCGCGGCTTCAAGGCGAGCGGACAGCTTCCATGATG 13851  
Qy 1981 CCCATTAGACACCTTCGATCTCGGATGATCATCTACACTCGGAGATCCACAGGTTG 2040  
Db 13852 CCCATTAGACACCTTCGATCTCGGATGATCATCTACACTCGGAGATCCACAGGTTG 13911  
Qy 2041 CCAAGGGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100  
Db 13912 CCAAGGGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13971  
Qy 2101 CGCTTGAATAGGCGCGGAGACAGGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160  
Db 13972 CGCTTGAATAGGCGCGGAGACAGGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14031  
Qy 2161 TCGGTCTATGATGTGTTGCGGATCTCTGGGCGGCGGAGTCAAGTGTGTCGAGCGG 2220  
Db 14032 TCGGTCTATGATGTGTTGCGGATCTCTGGGCGGCGGAGTCAAGTGTGTCGAGCGG 14091  
Qy 2221 TCAAGCTGCGGATCCGCGCATTTGGGAGAGTGTGATGAAAGAGAGAGTGAAGG 2280  
Db 14092 TCAAGCTGCGGATCCGCGCATTTGGGAGAGTGTGATGAAAGAGAGAGTGAAGG 14151  
Qy 2281 TGAAGCTGCGTCCGCGCTGATGCGGATGCTGTGAGCATTTTGAAGGTGCGCCGAT 2340  
Db 14152 TGAAGCTGCGTCCGCGCTGATGCGGATGCTGTGAGCATTTTGAAGGTGCGCCGAT 14211

2341 TCGCTGCTAGAGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGATCCCGTGGGCTG 2400  
14212 TCGCTGCTAGAGTCTCTGCGGCTTTCGCTGCTGAGCGGCGACTGATCCCGTGGGCTG 14271  
2401 CCTGCGAGGCTTCAGAGCCATCAGGCCCGGCGTGTGGTATGATAGCTTGGCGGGCCACC 2460  
14272 CTTGCGAGGCTTCAGAGCCATCAGGCCCGGCGTGTGGTATGATAGCTTGGCGGGCCACC 14331  
2461 GAAAGCTGATCTGATCTGATCGGATACCGGCGGAGAAAGTGAACCTTATCGTGGGCGAGC 2520  
14332 GAAAGCTGATCTGATCTGATCGGATACCGGCGGAGAAAGTGAACCTTATCGTGGGCGAGC 14391  
2521 ATCCCTACCGCGCTGCGCTGCGCAACAGAGCTTCAACGTCTGATGAGCGCTGCA 2580  
14392 ATCCCTACCGCGCTGCGCTGCGCAACAGAGCTTCAACGTCTGATGAGCGCTGCA 14451  
2581 CCGCGCCCGGCTTGGGTTCCGGGGCACTCTACATTTGGCGGGGTGGGCTGGCACTGGGC 2640  
14452 CCGCGCCCGGCTTGGGTTCCGGGGCACTCTACATTTGGCGGGGTGGGCTGGCACTGGGC 14511  
2641 TACTGCGCGATGAGAGAGAGAGCGCAAGAGCTTCTGTCACCCCGAGACGGGGAG 2700  
14512 TACTGCGCGATGAGAGAGAGAGCGCAAGAGCTTCTGTCACCCCGAGACGGGGAG 14571  
2701 CGCCTCTACAGAGCGGCGATCTGGGCGGCTACCTGCCGATGAGAAACATCGATTCAAG 2760  
14572 CGCCTCTACAGAGCGGCGATCTGGGCGGCTACCTGCCGATGAGAAACATCGATTCAAG 14631  
2761 GGGCGTGAAGACACCAATCAAGCTTGGCGGATACCGCGTTGAGCTGGGGAAATCGAG 2820  
14632 GGGCGTGAAGACACCAATCAAGCTTGGCGGATACCGCGTTGAGCTGGGGAAATCGAG 14691  
2821 GAAAGCTCAAGTGCATCCGAAAGTACGCGAGCGGATGATTGTCGCGTGGGAGAGAC 2880  
14692 GAAAGCTCAAGTGCATCCGAAAGTACGCGAGCGGATGATTGTCGCGTGGGAGAGAC 14751  
2881 GCGCGGAAACAAGCTCTTCTAGCTTATGTGTCCCGAGGGGCAACGAGACGCGCTCC 2940  
14752 GCGCGGAAACAAGCTCTTCTAGCTTATGTGTCCCGAGGGGCAACGAGACGCGCTCC 14811  
2941 GAGCGAGACGCGAGCTCTCAAGCGGAGGATCGACGCGAGAGCAACGCGCGGAGAGC 3000  
14812 GAGCGAGACGCGAGCTCTCAAGCGGAGGATCGACGCGAGAGCAACGCGCGGAGAGC 14871  
3001 GACGCTTGAAGCGAGCGGAGAGGATGAGTCAAGCTTCCGTCGACAGGACTCCGAGG 3060  
14872 GACGCTTGAAGCGAGCGGAGAGGATGAGTCAAGCTTCCGTCGACAGGACTCCGAGG 14931  
3061 GACCTGAGCGGAAAGCCGCTGCTGATCTGACCGGGCAGATCCGCGGAGCGGGGCTG 3120  
14932 GACCTGAGCGGAAAGCCGCTGCTGATCTGACCGGGCAGATCCGCGGAGCGGGGCTG 14991  
3121 GACGCTTGAAGCGGCTGCGCTGATGCGTCCGAAAGTCTTGAAGCGCCGATTCGTTGTT 3180  
14992 GACGCTTGAAGCGGCTGCGCTGATGCGTCCGAAAGTCTTGAAGCGCCGATTCGTTGTT 15051  
3181 GAGTTTGAATCTGATCTGAGCTGTTGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 3240  
15052 GAGTTTGAATCTGATCTGAGCTGTTGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 15111  
3241 AAATTCCTTATCTCATCTGCGGCGGAGCAAGTACCGCGTGAACCTAGCGTATCTCAA 3300  
15112 AAATTCCTTATCTCATCTGCGGCGGAGCAAGTACCGCGTGAACCTAGCGTATCTCAA 15171  
3301 TCCGCGCGGATGAGAGCGGCTGAGCAAGGAGCTTCTATTTATTAACCCGTTGAGCAAGCT 3360  
15172 TCCGCGCGGATGAGAGCGGCTGAGCAAGGAGCTTCTATTTATTAACCCGTTGAGCAAGCT 15231  
3361 TTGCTGAAGCTCTCGATCAAGGAGTCAAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 3420  
15232 TTGCTGAAGCTCTCGATCAAGGAGTCAAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 15291  
3421 GTGTTGATGAAGCGGCTTCAACCTCTGTTCTGTTGGGAGGATGAGCGGATCGAGTGC 3480

15292 GTGTTGATGAAGCGGCTTCAACCTCTGTTCTGTTGGGAGGATGAGCGGATCGAGTGC 15351  
3481 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3540  
15352 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15411  
3541 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3600  
15412 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 15471  
3601 GAAAGCTTGGCGGCTTCTGACCTGCGACATTCGAGCTTTACGTGACGAGCTG 3660  
15472 GAAAGCTTGGCGGCTTCTGACCTGCGACATTCGAGCTTTACGTGACGAGCTG 15531  
3661 GCGCGCGGAGTGAAGACCGCGGCAATTCAGGCTGTAAGCTGCGGAGAGATTCGCAAG 3720  
15532 GCGCGCGGAGTGAAGACCGCGGCAATTCAGGCTGTAAGCTGCGGAGAGATTCGCAAG 15591  
3721 AGCGCGCGAGAGCGCGGCGGCTTCCGCGCGAGAGCACTTCCGATATGCTT 3780  
15592 AGCGCGCGAGAGCGCGGCGGCTTCCGCGCGAGAGCACTTCCGATATGCTT 15651  
3781 CGGACCTTCTTGAAGACCAATCAACCGAGTACATGCTGCTTACATGCTTCTGAGGCTC 3840  
15652 CGGACCTTCTTGAAGACCAATCAACCGAGTACATGCTGCTTACATGCTTCTGAGGCTC 15711  
3841 GATGCTTGGCGGCTGAGCTGCAAGCGGAGGATGATGATGATGATGATGATGATGATGATGAT 3900  
15712 GATGCTTGGCGGCTGAGCTGCAAGCGGAGGATGATGATGATGATGATGATGATGATGATGAT 15771  
3901 GATGCTTGGCGGCTGAGCTGCAAGCGGAGGATGATGATGATGATGATGATGATGATGATGAT 3960  
15772 GATGCTTGGCGGCTGAGCTGCAAGCGGAGGATGATGATGATGATGATGATGATGATGATGAT 15831  
3961 CTCGTCGCGTCTGATCGGAGAGTCTCGGCTGAGAGTGTGCTGCGGCTTCACAGAGCTTC 4020  
15832 CTCGTCGCGTCTGATCGGAGAGTCTCGGCTGAGAGTGTGCTGCGGCTTCACAGAGCTTC 15891  
4021 GTGATCTTGGTGGAGACATGATTCATCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
15892 GTGATCTTGGTGGAGACATGATTCATCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15951  
4081 CTGAT 4140  
15952 CTGAT 16011  
4141 TCCGTTTGGCGGAGACTCGAGAGATCTAGATGATGATGATGATGATGATGATGATGATGAT 4200  
16012 TCCGTTTGGCGGAGACTCGAGAGATCTAGATGATGATGATGATGATGATGATGATGATGAT 16071  
4201 GAGGTTGGCGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4233  
16072 GAGGTTGGCGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16104  
RESULT 7  
AR208671  
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 1 from patent US 6383787.  
ACCESSION AR208671  
VERSION AR208671.1 GI:21509886  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)  
AUTHORS Schupp,T., Iygon,J.,Madison., Molnar,I., Zikkle,R., Cyr,D.,Dawn. and  
Gorlach,J.  
TITLE Genes for the biosynthesis of epoethilones  
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;  
FEATURES  
source 1..68750

/organism="unknown"  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN  
Query Match 100.0%; Score 4233; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ATGAGCATCAATCAGCTTCTGAAAGAGCTCGAGCACAGGGTATCAAGTGGCGCGCAT 60  
11872 ATGAGCATCAATCAGCTTCTGAAAGAGCTCGAGCACAGGGTATCAAGTGGCGCGCAT 11931  
61 GGGGAGCGCTCCAGATACAGGCGCCCAAGAACGCCCTGAACCCGACCTGCTCGTCA 120  
11932 GGGGAGCGCTCCAGATACAGGCGCCCAAGAACGCCCTGAACCCGACCTGCTCGTCA 11991  
121 ATCTCGAGACAAAAGACAGATCTCGAGATGCTCCGTCAAGAGATCCCGCAGATCC 180  
11992 ATCTCGAGACAAAAGACAGATCTCGAGATGCTCCGTCAAGAGATCCCGCAGATCC 12051  
181 ATCTGCCCCCGCCAGCGGAGCGGACGCTCCGTTCTCTCAAGACATCCAGATCC 240  
12052 ATCTGCCCCCGCCAGCGGAGCGGACGCTCCGTTCTCTCAAGACATCCAGATCC 12111  
241 TACTGCTGGGCGGACAGAGGCTTTAGGTCGCCAGGGGATCCAGGCTATGCGGA 300  
12112 TACTGCTGGGCGGACAGAGGCTTTAGGTCGCCAGGGGATCCAGGCTATGCGGA 12171  
301 TACGACTGACGAGATCTCGACGTGCGAGGCTGAGCGCGCTTTGGAAGTCTGCGG 360  
12172 TACGACTGACGAGATCTCGACGTGCGAGGCTGAGCGCGCTTTGGAAGTCTGCGG 12231  
361 CGGACGACATGCTTGGGCGCCACACGCTGCCGACATATGAGGTATCGAGCTTAA 420  
12232 CGGACGACATGCTTGGGCGCCACACGCTGCCGACATATGAGGTATCGAGCTTAA 12291  
421 GTGAGCGCGGACATGAGATCATGATCTGCGCGGCTGACCGGAGCACAGGGAAAGG 480  
12292 GTGAGCGCGGACATGAGATCATGATCTGCGCGGCTGACCGGAGCACAGGGAAAGG 12351  
481 AGGCTGCTGCTGCGAGATGAGATGTCGACCGGATATGACACCGAGCGCTCCG 540  
12352 AGGCTGCTGCTGCGAGATGAGATGTCGACCGGATATGACACCGAGCGCTCCG 12411  
541 CTCTATCAGCTGCTGCGGCTGGGCTGAGACAGCGGCAAAACCGCTTCTGCTCAGTATC 600  
12412 CTCTATCAGCTGCTGCGGCTGGGCTGAGACAGCGGCAAAACCGCTTCTGCTCAGTATC 12471  
601 GATTCATTAAGTTGACCTAGGACGCTGTCATCATCTTCAAGGACTGGGCTCAGCTTC 660  
12472 GATTCATTAAGTTGACCTAGGACGCTGTCATCATCTTCAAGGACTGGGCTCAGCTTC 12531  
661 TACGAAGATCCGAGACCTTCTCCCTGCTCTGAGACTCTGTAACCGCATTTATGATTC 720  
12532 TACGAAGATCCGAGACCTTCTCCCTGCTCTGAGACTCTGTAACCGCATTTATGATTC 12591  
721 GCGCTGGAATCTCGCAAGAGTCTGAGCGGCTCAAGATCGATGATTTACTGGAAGCGG 780  
12592 GCGCTGGAATCTCGCAAGAGTCTGAGCGGCTCAAGATCGATGATTTACTGGAAGCGG 12651  
781 CGCATGCGGAGTCCACCTCGCGGAGCGCTTCGATGAAGGCGGATCATCTACCTTG 840  
12652 CGCATGCGGAGTCCACCTCGCGGAGCGCTTCGATGAAGGCGGATCATCTACCTTG 12711  
841 AAGAGATCCGCTTCCGAGACAGAGCATGAGCTGCGCTGCACTTCTGAGGTCGATG 900  
12712 AAGAGATCCGCTTCCGAGACAGAGCATGAGCTGCGCTGCACTTCTGAGGTCGATG 12771  
901 AAGCGCGTGTGCGGAGAGCGCGGCTGACCCCGAGCGGCTCATCTGCTGATTTTC 960  
12772 AAGCGCGTGTGCGGAGAGCGCGGCTGACCCCGAGCGGCTCATCTGCTGATTTTC 12831  
961 GAGGTATGCGGCGCTGAGAGCGGAGCCCGGTTTACGCTCAACATTAAGCTTTCAAC 1020

12832 GAGGTATGCGGCGCTGAGAGCGGAGCCCGGTTTACGCTCAACATTAAGCTTTCAAC 12891  
1021 CGGCTCCCGCTCATCCGCGGCTGAAAGATATCACCGGAGACTTCAAGTGTCTCTC 1080  
12892 CGGCTCCCGCTCATCCGCGGCTGAAAGATATCACCGGAGACTTCAAGTGTCTCTC 12951  
1081 CTGGAATCTGACACCACTCGGAGCAAGAGCTTCGAACAGCGGCTTAAGCTTCAAG 1140  
12952 CTGGAATCTGACACCACTCGGAGCAAGAGCTTCGAACAGCGGCTTAAGCTTCAAG 13011  
1141 CAGCTGGGAAAGCATGATATCACTGCAAGCTTAAGCTTCAAGCTCAGAGAGGCC 1200  
13012 CAGCTGGGAAAGCATGATATCACTGCAAGCTTAAGCTTCAAGCTCAGAGAGGCC 13071  
1201 GCCCGGATCTTGGGATCCAAAGAGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTT 1260  
13072 GCCCGGATCTTGGGATCCAAAGAGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTT 13131  
1261 AACGAGCAAGTCTGTGTGCTCACTCTGTTGCAAGAGCTCGGAACCTCGGTTACACAGC 1320  
13132 AACGAGCAAGTCTGTGTGCTCACTCTGTTGCAAGAGCTCGGAACCTCGGTTACACAGC 13191  
1321 ACGGACTCTCTCAGCTGCTGATCATCAGCTCTACAGACAGATGGGAGCTCTGTC 1380  
13192 ACGGACTCTCTCAGCTGCTGATCATCAGCTCTACAGACAGATGGGAGCTCTGTC 13251  
1381 CTGCGGTGGACATGCTGACAGAGTGTCCCGCCGACCTTTGAGACATGCTCGAA 1440  
13252 CTGCGGTGGACATGCTGACAGAGTGTCCCGCCGACCTTTGAGACATGCTCGAA 13311  
1441 GGTATGCTGTTTTTCTCCGCGCTCACTAGGAACATAGGGGGAACAGTGGCTGT 1500  
13312 GGTATGCTGTTTTTCTCCGCGCTCACTAGGAACATAGGGGGAACAGTGGCTGT 13371  
1501 TCGCTTCGCGCTGCCAGCTAGAAAGCGGAGGAGCGCAACGCGATGCTG 1560  
13372 TCGCTTCGCGCTGCCAGCTAGAAAGCGGAGGAGCGCAACGCGATGCTG 13431  
1561 AGCGACATACGCTGACAGGCTGTTGCGGCGGCTGAGACAGCTGCCATGACGCTC 1620  
13432 AGCGACATACGCTGACAGGCTGTTGCGGCGGCTGAGACAGCTGCCATGACGCTC 13491  
1621 GCCGTGTGTGCGGCGGAAAGCGCTCAAGCAAGAGCTTTCGCGCTTCCGCGGGA 1680  
13492 GCCGTGTGTGCGGCGGAAAGCGCTCAAGCAAGAGCTTTCGCGCTTCCGCGGGA 13551  
1681 CTTCGCGCGGCTCGGAGAGGAGGACGCGCGAACAATGTCGCGGTGATG 1740  
13552 CTTCGCGCGGCTCGGAGAGGAGGACGCGCGAACAATGTCGCGGTGATG 13611  
1741 GAGAAAGCTGGAGACAGGTTGTCCGGTTCTTCGCGTGTCTGAGTCAGCGCGCTAC 1800  
13612 GAGAAAGCTGGAGACAGGTTGTCCGGTTCTTCGCGTGTCTGAGTCAGCGCGCTAC 13671  
1801 GTGCCGATCGATGCGGACCTACCGGAGAGGATCTCACTCTCTGATCATGATG 1860  
13672 GTGCCGATCGATGCGGACCTACCGGAGAGGATCTCACTCTCTGATCATGATG 13731  
1861 GTAAAGCTGCTGCTGACAGCAGCATGCTGATGCAACTGTCACTGCGCGGAGTC 1920  
13732 GTAAAGCTGCTGCTGACAGCAGCATGCTGATGCAACTGTCACTGCGCGGAGTC 13791  
1921 CAGCGGCTGCTGCTGAGCGAGGCGCGCTCGAAGGCAAGGCAACAGCTTCGATGATG 1980  
13792 CAGCGGCTGCTGCTGAGCGAGGCGCGCTCGAAGGCAAGGCAACAGCTTCGATGATG 13851  
1981 CCCATTTGAGACCTTTCGATCTCGCGTATGCTCATCACTCGGAGTCCACAGGTTG 2040  
13852 CCCATTTGAGACCTTTCGATCTCGCGTATGCTCATCACTCGGAGTCCACAGGTTG 13911  
2041 CCCAAGGGGATGATGATCATCGGGGTGCGCTCAACACCATCTGAGACATCAAGG 2100

Db 13912 CCCAAGGGGGTGAATATGCATATCGGGGTCCGCTCAACACCATCTCGACATCAACGAG 13971  
Qy 2101 CGCTTGAATAATGAGCCCGGAGACAGAGGTGTGCGCTCTCTCGCTGAGTTCCATCTC 2160  
Db 13972 CGCTTGAATAATGAGCCCGGAGACAGAGGTGTGCGCTCTCTCGCTGAGTTCCATCTC 14031  
Qy 2161 TCGGTCTAATGATGTGTGGGATCTCTGCGCGCGGGCGGTACGATCTGTGTGCGGACGCG 2220  
Db 14032 TCGGTCTAATGATGTGTGGGATCTCTGCGCGCGGGCGGTACGATCTGTGTGCGGACGCG 14091  
Qy 2221 TCCAGAGTCCCGGATCCCGGCGGATTTGGGAGAGTTGATTCGAACGAGAAAGGTGACGCTG 2280  
Db 14092 TCCAGAGTCCCGGATCCCGGCGGATTTGGGAGAGTTGATTCGAACGAGAAAGGTGACGCTG 14151  
Qy 2281 TGAAGTCCGCTCCGAGCTGATGCGGATGCTGTCGAGCATTTTGAAGGCTGCCCGCAT 2340  
Db 14152 TGAAGTCCGCTCCGAGCTGATGCGGATGCTGTCGAGCATTTTGAAGGCTGCCCGCAT 14211  
Qy 2341 TCGCTCGCTAGGTCTCTGCGGCTTTTCTGCTGAGCGGCGCATGGAATCCCGTGGGCTTG 2400  
Db 14212 TCGCTCGCTAGGTCTCTGCGGCTTTTCTGCTGAGCGGCGCATGGAATCCCGTGGGCTTG 14271  
Qy 2401 CCTGCGAGCTCCAGGCGCATCAAGCCCGGCGCTGTGCTGATATAGCTTGGGCGGCGCAC 2460  
Db 14272 CCTGCGAGCTCCAGGCGCATCAAGCCCGGCGCTGTGCTGATATAGCTTGGGCGGCGCAC 14331  
Qy 2461 GAAAGCTCGATCTGGTCCATCTGGGTACCCCGTGAAGAAAGTGCATCTGCGGCGAGC 2520  
Db 14332 GAAAGCTCGATCTGGTCCATCTGGGTACCCCGTGAAGAAAGTGCATCTGCGGCGAGC 14391  
Qy 2521 ATCCCTCAACGCGCTCTGCTGCGCAACGAGCTTCAAGTCTGATGAGCGCTCGA 2580  
Db 14392 ATCCCTCAACGCGCTCTGCTGCGCAACGAGCTTCAAGTCTGATGAGCGCTCGA 14451  
Qy 2581 CCGGCGCGGCTGTGGGTTCCGGGCGAATCTTACATTTGGCGGGGTGGGCTGGCACTGGG 2640  
Db 14452 CCGGCGCGGCTGTGGGTTCCGGGCGAATCTTACATTTGGCGGGGTGGGCTGGCACTGGG 14511  
Qy 2641 TACTGGCGGATGAG 2700  
Db 14512 TACTGGCGGATGAG 14571  
Qy 2701 CGCTCTCAACAGACCGGCGATCTGGGCGCTTACCTTCCCGATGAGAAACATCGAGTTCA 2760  
Db 14572 CGCTCTCAACAGACCGGCGATCTGGGCGCTTACCTTCCCGATGAGAAACATCGAGTTCA 14631  
Qy 2761 GGGCGTGAAGACCAACCAATCAAGCTTGGCGGATACCGGCTTGAAGTGGGAAATCGAG 2820  
Db 14632 GGGCGTGAAGACCAACCAATCAAGCTTGGCGGATACCGGCTTGAAGTGGGAAATCGAG 14691  
Qy 2821 GAAACGCTCAAGTGCATCCGAAAGTACGCGAGCGGAGTGTGCGCGTGGGAAACGAC 2880  
Db 14692 GAAACGCTCAAGTGCATCCGAAAGTACGCGAGCGGAGTGTGCGCGTGGGAAACGAC 14751  
Qy 2881 GCGGCGAACAAGCTCTTCTAGCTTATGTGTCCCGAGGAGACACGAGAGCGCTGCC 2940  
Db 14752 GCGGCGAACAAGCTCTTCTAGCTTATGTGTCCCGAGGAGACACGAGAGCGCTGCC 14811  
Qy 2941 GAGCAGAGACGAGCTCTCAAGACCGAGCGGATTCGACGCGAGACACGCGCGCAAGCG 3000  
Db 14812 GAGCAGAGACGAGCTCTCAAGACCGAGCGGATTCGACGCGAGACACGCGCGCAAGCG 14871  
Qy 3001 GACGCGTGAAGCGAGCGGAGAGAGAGTTCAGAGCTGCGTCAACGAGATCCCGGAGG 3060  
Db 14872 GACGCGTGAAGCGAGCGGAGAGAGAGTTCAGAGCTGCGTCAACGAGATCCCGGAGG 14931  
Qy 3061 GACCTGAACGAGACCGCTGCTGATTCGACCGGCGAGATCCGCGGAGCGGCGCTG 3120  
Db 14932 GACCTGAACGAGACCGCTGCTGATTCGACCGGCGAGATCCGCGGAGCGGCGCTG 14991  
Qy 3121 GACGCTTACGCGCGTTCGCTGAGCTTCGGAACGTTCTTGAAGCGCGGATTCGTTGTT 3180  
Db 14992 GACGCTTACGCGCGTTCGCTGAGCTTCGGAACGTTCTTGAAGCGCGGATTCGTTGTT 15051

Qy 3181 GAGTTTGTGATTCCTGAGCTGTGAGAGCGTGAAGCCGACGCGCGACCTTCC 3240  
Db 15052 GAGTTTGTGATTCCTGAGCTGTGAGAGCGTGAAGCCGACGCGCGACCTTCC 15111  
Qy 3241 AAATTCGTTATCCATCGCGCGGACGAGTACCCTGTCAACTTACGCTATGTCAA 3300  
Db 15112 AAATTCGTTATCCATCGCGCGGACGAGTACCCTGTCAACTTACGCTATGTCAA 15171  
Qy 3301 TCCGCGCATCGAGGGGCTGAGAGAGGCTTATATTAATTCACCCGTTGAGACCGT 3360  
Db 15172 TCCGCGCATCGAGGGGCTGAGAGAGGCTTATATTAATTCACCCGTTGAGACCGT 15231  
Qy 3361 TTGCTGAAGCTCTCGATCAAGGATTCGAGCGGAGCGACGTTGCGCAAACTTCGAC 3420  
Db 15232 TTGCTGAAGCTCTCGATCAAGGATTCGAGCGGAGCGACGTTGCGCAAACTTCGAC 15291  
Qy 3421 GTTTCGATGAAGCGGCTTCAACTCTGTTCTGTTGCGGAGAGTGAAGCCATCGAGTGC 3480  
Db 15292 GTTTCGATGAAGCGGCTTCAACTCTGTTCTGTTGCGGAGAGTGAAGCCATCGAGTGC 15351  
Qy 3481 CTGATGAATGCTGTCGCGAGAAATTTTCCGCTGAGAGCGGATATATGCGGAGCTC 3540  
Db 15352 CTGATGAATGCTGTCGCGAGAAATTTTCCGCTGAGAGCGGATATATGCGGAGCTC 15411  
Qy 3541 CTGATGAAGAGCGGCTCTCTGCAACATCGGCGCTCTGCGTGGGCAATTCATTTT 3600  
Db 15412 CTGATGAAGAGCGGCTCTCTGCAACATCGGCGCTCTGCGTGGGCAATTCATTTT 15471  
Qy 3601 GAAACGTTTCGCGCGGTTCTGACCTGCGACATTTGAGCGTTTACGTGACGCGATGCTG 3660  
Db 15472 GAAACGTTTCGCGCGGTTCTGACCTGCGACATTTGAGCGTTTACGTGACGCGATGCTG 15531  
Qy 3661 GCGCGCGGCTGAGACCGCGGCAATTCGAGGCTGTAAGCTCGGCAAGATTCCTCACCG 3720  
Db 15532 GCGCGCGGCTGAGACCGCGGCAATTCGAGGCTGTAAGCTCGGCAAGATTCCTCACCG 15591  
Qy 3721 AGCGCGCGAGAGCGCGGCGGCTTCCGCGCGGAGAGAGAGAGAGAGAGAGAGAG 3780  
Db 15592 AGCGCGCGAGAGCGCGGCGGCTTCCGCGCGGAGAGAGAGAGAGAGAGAGAGAG 15651  
Qy 3781 CGGACCTTCTTGAAGACCAATCAACCGAGTATGCTGCTTACGCTTCTGAGAGCTC 3840  
Db 15652 CGGACCTTCTTGAAGACCAATCAACCGAGTATGCTGCTTACGCTTCTGAGAGCTC 15711  
Qy 3841 GATGCTTTCGCGCTGACCTTCAACGCGAGGTTGATGCTTACGCTTTCGCGAGCGGAG 3900  
Db 15712 GATGCTTTCGCGCTGACCTTCAACGCGAGGTTGATGCTTACGCTTTCGCGAGCGGAG 15771  
Qy 3901 GATACCTGCGCGCGGCAATTCGAGGAGACAGCGGAGAGAGAGAGAGAGAGAGAG 3960  
Db 15772 GATACCTGCGCGCGGCAATTCGAGGAGACAGCGGAGAGAGAGAGAGAGAGAGAG 15831  
Qy 3961 CTCGTCGCGCTGATACGAGAGTCTCGGCGTGAAGGTGTGCGGCTTCAGACAGCTTC 4020  
Db 15832 CTCGTCGCGCTGATACGAGAGTCTCGGCGTGAAGGTGTGCGGCTTCAGACAGCTTC 15891  
Qy 4021 GTCGATCTTGGTGGACATTCATTCATCTGTTGAGTGAAGAGCTTTCAGAAAGG 4080  
Db 15892 GTCGATCTTGGTGGACATTCATTCATCTGTTGAGTGAAGAGCTTTCAGAAAGG 15951  
Qy 4081 CTGATGAGAGATTCGACATCAACGAGTGTTCAGAGTCCCGAAGCTTGGCTGCTGCG 4140  
Db 15952 CTGATGAGAGATTCGACATCAACGAGTGTTCAGAGTCCCGAAGCTTGGCTGCTGCG 16011  
Qy 4141 TCCGTTTTCGCGAGACTCGAGAGATTCAGAGCGGCGGAAATGACAGACCGAGTG 4200  
Db 16012 TCCGTTTTCGCGAGACTCGAGAGATTCAGAGCGGCGGAAATGACAGACCGAGTG 16071  
Qy 4201 GAGGTTTCGCGAG 4233  
Db 16072 GAGGTTTCGCGAG 16104



```

RESULT 8
AF217189
LOCUS       58733 bp      DNA              linear      BCT 09-JUN-2000
DEFINITION  Sorangium cellulosum putative transposase gene, partial cds;
              putative transposase gene, complete cds; epothilone biosynthesis
              gene cluster, complete sequence; putative membrane protein gene,
              complete cds.
ACCESSION   AF217189
VERSION     AF217189.1  GI:7453554
KEYWORDS
SOURCE      Polyangium cellulosum
            Polyangium cellulosum
            Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
            Soranginae; Polyangiaceae; Polyangium.
REFERENCE   1 (bases 1 to 58733)
            Julien,B., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
            Cloning and heterologous expression of the epothilone gene cluster
            Science 287 (4543), 640-642 (2000)
JOURNAL     MEDLINE
            PUBMED 20115953
            PUBMED 10649995
REFERENCE   2 (bases 1 to 58733)
            Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
            Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
            Gene 249 (1-2), 153-160 (2000)
JOURNAL     MEDLINE
            PUBMED 20293058
            PUBMED 10831849
REFERENCE   3 (bases 1 to 58733)
            Julien,B.
            Direct Submmission
            Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES
    source   1..58733
             /organism="Polyangium cellulosum"
             /mol_type="genomic DNA"
             /strain="SMP44"
             /db_xref="taxon:56"
             <1..992
             /note="orfA"
             /codon_start=3
             /transl_table=1
             /product="putative transposase"
             /protein_id="AAF62878.1"
             /db_xref="GI:7453555"
             /transl_table="VRGHEVAFADPGVRYLYLDNLKNAVERHGDAIRPHPTLLAL
             SADRPFEPRVAIVAGNGEKVERAIRVYRGFFEARAYADLGLNQATQWTSALAL
             DRSWEDBARATVROAFDERSVILRHPTPPDHERVEVEGKTPYARFDLNDYSVPH
             DRTATVETVLADLSQVRLADNGNQIVATVTRMSDGGQIQCEPHLQRLVDEKRRAREHR
             GLDVARARSSSOAFRLTIVARGNVCSAIRLRLQILDVAGAALEALVLEVRDITI
             HIGAVROVIDRRREBRHLPVPVSLPVRGHEALVYVPHSLITTDALKQDPT"
             989..1501
             /note="orfB"
             /codon_start=1
             /transl_table=1
             /product="putative transposase"
             /protein_id="AAF62879.1"
             /db_xref="GI:7453556"
             /transl_table="MTDITPTETKRLKSLGLFGLLACWEQLADKPMILREVLAIEER
             RHKSLERLRLKNSRVAAFKPMTDPDSSWPKKIDIEAVDDLSDRSADLLPEVVRRDV
             AOKLLSLSTNKAFADMGQVFFHAACVTLVDRLVLRABVIEIEAESYRLKEAKENLATN
             RTKORRTKKH"
             1998..6263
             /gene="epoA"
             /note="epoA"
             1998..6263
             /gene="epoA"
             /note="Type I polyketide synthase; epothilone PKS loading
             module"
             /codon_start=1
             /transl_table=1
             /product="epoA"

```

gene  
CDS

protein\_id="AA062880.1"  
/db\_xref="gi:7453557"  
/translation="MADRPTEERAEDPPIAIVGAGCRLPGGVINDLSGPTVLTEGSRDVC  
GQVPERDADAAMFDPDLDAPEGKTPVTRAFVLSGVACFDPSFFPGISPRELADMPDR  
LLLEVCWALEENAAIAPSAIVGTETFTFQIGSEVEPALPRTATASEIDIAHGGLGM  
PSVAGRISVYLGLRGPCVAVDTRAYSLSIVAHVLAQSLSGSGCSCTALAGVSLSP  
STVLWSTRTALATDGRCAKFAFSAEDFGREGCAVVVLRSLGSAADGRIALAVLRG  
SAIHNDGSSGLVAVNGSSSOEIVLKRALADAGCASSGVVEAHGTTGLDPIETIOA  
LNAVYAGRDVATPLILGSVKTNLGHVEVSAITGLIKVYLSTIOHQIPIHLIAOLN  
PRISMGDLRLVTYARTPTPMDNTPRAGISGSGMSTNAHVLEEARPTCTPPAPE  
RPAELVIAATTAALDAHAARLDHLETYSQCLGDVAPSLTTSAMEHRLAVAT  
SSEGRALDIAAAGQTPGVVGRRIADSSSGKLAFETGGAGTQLMGRLXVMPAR  
REAPLCYRLNQCGLDRLPEVMAEASVDALADOTATOPALFTFEVALTAALRS  
MGVPELVAGHSIGELVAACVAGVFEPIVLAARGLMOLPGGAAVSLAADEA  
DVAANAAPHASVSLAANGPDQVVIAGACQPHALIAAANAAGARTKALHVSNAHS  
PLMAPLEAFGRVAVESVYRPSSTIVASNLISGAGDEVSFGVWRHAEEVFRPD  
VKALHAAGACFVVEGPKSTLGLVAPCLDPADPALAASSRGRDEPVALTEALGLM  
AVGSLVAGSRFPGSGRRVPLPTPMQRERYMIDTKAADARDBRAFGAHGEBVKG  
GAVGGDRSRALDHPPEPSGRREKVEAADRPRLIEDPGVLDRLVATREBRAG  
LGEVRIANDAGSLSPNDVALYVPPDLGCKNPMLLIGCEAGGAIIVANGEDVRIQ  
VGQVILASGACFATHTTTSALVYPPDLASATEAAMVATITMYALDGIARQV  
GERVLIHAATGVGLAAVQMAQHVGAHVHATAGPEKRAVLESLGVRVYSDSDRIV  
ADVATMGEGCEVDVNLISGELIDKSNLRSHRGVEGLCKCVAADQALGRLPRL  
NLNSELVDLRGMLERPARVALFEEELGSIAGVFPPEIATPLPVAADARSMAQ  
AOKGLVLTIVGDEVOIRIPTHAAGSPGEGDRLDRADGAPAAARAALEAFRLQ  
VSQVLRPEIIVGAEALPTRLGMDSLMAVLENRIRASLKLKLTSTPLSFIALLT  
QNLIDLATLISLERRVAENDRAGVSDPVSAGDDMEITAL"

gene  
CDS

protein\_id="AA062881.1"  
/db\_xref="gi:7453558"  
/translation="MTINQLNELHEHGVYLAADGERLQIQAQRNALNPULARISEH  
KSTILTMLROKLPAESIVPAERHVPFPLTDIGSVWLGRTAALTPVPSGIAVREYD  
CTDLDVALASFAKRVVAAHDMLEAHLPMQVOVIEBKWDADIEIDLRLDSTSEA  
RLVSDRMSHRIRYDTERPPLVHVAARLDEOGRVLVSLDLINVLGSGSIIFKML  
SFYDPEPSTLVELSLRTPVLALESKEAHORSMVDKRVVLEPPMPAMD  
PSTIREIRFRTEQWMLPDSKSRKQGVGRGLTPGVILTAASVYIGRSSAPRPLT  
NITFNRLPVHPRVNDITGDFTSNVLDITTDKSSBQAKRIQELREAMDHCVS  
GIEVQREAKRVLGIQRGALFPVVLTSALNQVVGVSLORLGTPVYTSQTPOLLDH  
OLYHGDGVLVAMDIIVGVPEPDLDDMLEAAVAFRLRLTEBSEMSGRSLPDALE  
ARASINENLSLSEHTLGLFAARVEOLOPMLVAVNSARKUTLYEELSRSRGALYLR  
EOGARPNITVAVVEKGMEOVAVNTIAVLESGAAVYVDLADLPAERLHYILLDEHEVLA  
LTPMLDCKLSWPGIORLRLVSDAGUGDDQPLMPPIQTSPDLAVIYTTSGSGPK  
GVMTIDHRAAVNTLIDINERPELIGDGVLAALSISPELSTYVDVFGLLAAGQTLVDDA  
SKLDDPAMMALIEREKVYTWNSPVALMRLVHSBGRDPSLARSLTLSSDMMIPV  
GLPGSLOLRGVAVSILSGATEASLWSISGVPANVDPWASLIPYGRPLANQTFHVL  
BALBRPWPVNGOLYIGVGVLGALGVEDEKTRNSLIVHEGTENRILYKTDGLGRVLPD  
GNIFMRGENDQITLRGVRVBLGIEETLKSHNVDAVTVIPEGNAANKLLIAVVP  
EGTRRAAEQDASLTERVDARAHAAADGSLSGERYQFLAHHGRLRDDCKRPVVDL  
TGLVREAGLGVAVARRSVRFLFEPALPEVEFRGLSCSSPDGDAALPEKPYPSAG  
STYVQVTAAGVAKSGRIGVDEGFPYVHPFHRILKLYSDHIEBGAHPQNFVFDAA  
FGLLVFGIDILIESLGSRSRECLFAGMAQOLMEQASNCIQCVPQGFQEVOR  
PVLIDRSHSDVYHGMGLGVRDPFOVCTLGDGDSPPARATTRGAAPRGDHPDILRD  
FLRKLTPRYWPTVYFVELDALPLTNSKQVVRKALARKKOTSSRRHSGHTAPRALREI  
LVAAVREVLGLVEVGLQOVSFVDEAGTISITHVRRKSLQKRLDREIATTELFOYPLNGS  
LASGARRRSKDLERPMQDRVEARRRRRS"

gene  
CDS

protein\_id="AA062882.1"  
/db\_xref="gi:7453559"  
/translation="MADRPTEERAEDPPIAIVGAGCRLPGGVINDLSGPTVLTEGSRDVC  
GQVPERDADAAMFDPDLDAPEGKTPVTRAFVLSGVACFDPSFFPGISPRELADMPDR  
LLLEVCWALEENAAIAPSAIVGTETFTFQIGSEVEPALPRTATASEIDIAHGGLGM  
PSVAGRISVYLGLRGPCVAVDTRAYSLSIVAHVLAQSLSGSGCSCTALAGVSLSP  
STVLWSTRTALATDGRCAKFAFSAEDFGREGCAVVVLRSLGSAADGRIALAVLRG  
SAIHNDGSSGLVAVNGSSSOEIVLKRALADAGCASSGVVEAHGTTGLDPIETIOA  
LNAVYAGRDVATPLILGSVKTNLGHVEVSAITGLIKVYLSTIOHQIPIHLIAOLN  
PRISMGDLRLVTYARTPTPMDNTPRAGISGSGMSTNAHVLEEARPTCTPPAPE  
RPAELVIAATTAALDAHAARLDHLETYSQCLGDVAPSLTTSAMEHRLAVAT  
SSEGRALDIAAAGQTPGVVGRRIADSSSGKLAFETGGAGTQLMGRLXVMPAR  
REAPLCYRLNQCGLDRLPEVMAEASVDALADOTATOPALFTFEVALTAALRS  
MGVPELVAGHSIGELVAACVAGVFEPIVLAARGLMOLPGGAAVSLAADEA  
DVAANAAPHASVSLAANGPDQVVIAGACQPHALIAAANAAGARTKALHVSNAHS  
PLMAPLEAFGRVAVESVYRPSSTIVASNLISGAGDEVSFGVWRHAEEVFRPD  
VKALHAAGACFVVEGPKSTLGLVAPCLDPADPALAASSRGRDEPVALTEALGLM  
AVGSLVAGSRFPGSGRRVPLPTPMQRERYMIDTKAADARDBRAFGAHGEBVKG  
GAVGGDRSRALDHPPEPSGRREKVEAADRPRLIEDPGVLDRLVATREBRAG  
LGEVRIANDAGSLSPNDVALYVPPDLGCKNPMLLIGCEAGGAIIVANGEDVRIQ  
VGQVILASGACFATHTTTSALVYPPDLASATEAAMVATITMYALDGIARQV  
GERVLIHAATGVGLAAVQMAQHVGAHVHATAGPEKRAVLESLGVRVYSDSDRIV  
ADVATMGEGCEVDVNLISGELIDKSNLRSHRGVEGLCKCVAADQALGRLPRL  
NLNSELVDLRGMLERPARVALFEEELGSIAGVFPPEIATPLPVAADARSMAQ  
AOKGLVLTIVGDEVOIRIPTHAAGSPGEGDRLDRADGAPAAARAALEAFRLQ  
VSQVLRPEIIVGAEALPTRLGMDSLMAVLENRIRASLKLKLTSTPLSFIALLT  
QNLIDLATLISLERRVAENDRAGVSDPVSAGDDMEITAL"

gene  
CDS

protein\_id="AA062883.1"  
/db\_xref="gi:7453560"  
/translation="MADRPTEERAEDPPIAIVGAGCRLPGGVINDLSGPTVLTEGSRDVC  
GQVPERDADAAMFDPDLDAPEGKTPVTRAFVLSGVACFDPSFFPGISPRELADMPDR  
LLLEVCWALEENAAIAPSAIVGTETFTFQIGSEVEPALPRTATASEIDIAHGGLGM  
PSVAGRISVYLGLRGPCVAVDTRAYSLSIVAHVLAQSLSGSGCSCTALAGVSLSP  
STVLWSTRTALATDGRCAKFAFSAEDFGREGCAVVVLRSLGSAADGRIALAVLRG  
SAIHNDGSSGLVAVNGSSSOEIVLKRALADAGCASSGVVEAHGTTGLDPIETIOA  
LNAVYAGRDVATPLILGSVKTNLGHVEVSAITGLIKVYLSTIOHQIPIHLIAOLN  
PRISMGDLRLVTYARTPTPMDNTPRAGISGSGMSTNAHVLEEARPTCTPPAPE  
RPAELVIAATTAALDAHAARLDHLETYSQCLGDVAPSLTTSAMEHRLAVAT  
SSEGRALDIAAAGQTPGVVGRRIADSSSGKLAFETGGAGTQLMGRLXVMPAR  
REAPLCYRLNQCGLDRLPEVMAEASVDALADOTATOPALFTFEVALTAALRS  
MGVPELVAGHSIGELVAACVAGVFEPIVLAARGLMOLPGGAAVSLAADEA  
DVAANAAPHASVSLAANGPDQVVIAGACQPHALIAAANAAGARTKALHVSNAHS  
PLMAPLEAFGRVAVESVYRPSSTIVASNLISGAGDEVSFGVWRHAEEVFRPD  
VKALHAAGACFVVEGPKSTLGLVAPCLDPADPALAASSRGRDEPVALTEALGLM  
AVGSLVAGSRFPGSGRRVPLPTPMQRERYMIDTKAADARDBRAFGAHGEBVKG  
GAVGGDRSRALDHPPEPSGRREKVEAADRPRLIEDPGVLDRLVATREBRAG  
LGEVRIANDAGSLSPNDVALYVPPDLGCKNPMLLIGCEAGGAIIVANGEDVRIQ  
VGQVILASGACFATHTTTSALVYPPDLASATEAAMVATITMYALDGIARQV  
GERVLIHAATGVGLAAVQMAQHVGAHVHATAGPEKRAVLESLGVRVYSDSDRIV  
ADVATMGEGCEVDVNLISGELIDKSNLRSHRGVEGLCKCVAADQALGRLPRL  
NLNSELVDLRGMLERPARVALFEEELGSIAGVFPPEIATPLPVAADARSMAQ  
AOKGLVLTIVGDEVOIRIPTHAAGSPGEGDRLDRADGAPAAARAALEAFRLQ  
VSQVLRPEIIVGAEALPTRLGMDSLMAVLENRIRASLKLKLTSTPLSFIALLT  
QNLIDLATLISLERRVAENDRAGVSDPVSAGDDMEITAL"

gene  
CDS

protein\_id="AA062884.1"  
/db\_xref="gi:7453561"  
/translation="MADRPTEERAEDPPIAIVGAGCRLPGGVINDLSGPTVLTEGSRDVC  
GQVPERDADAAMFDPDLDAPEGKTPVTRAFVLSGVACFDPSFFPGISPRELADMPDR  
LLLEVCWALEENAAIAPSAIVGTETFTFQIGSEVEPALPRTATASEIDIAHGGLGM  
PSVAGRISVYLGLRGPCVAVDTRAYSLSIVAHVLAQSLSGSGCSCTALAGVSLSP  
STVLWSTRTALATDGRCAKFAFSAEDFGREGCAVVVLRSLGSAADGRIALAVLRG  
SAIHNDGSSGLVAVNGSSSOEIVLKRALADAGCASSGVVEAHGTTGLDPIETIOA  
LNAVYAGRDVATPLILGSVKTNLGHVEVSAITGLIKVYLSTIOHQIPIHLIAOLN  
PRISMGDLRLVTYARTPTPMDNTPRAGISGSGMSTNAHVLEEARPTCTPPAPE  
RPAELVIAATTAALDAHAARLDHLETYSQCLGDVAPSLTTSAMEHRLAVAT  
SSEGRALDIAAAGQTPGVVGRRIADSSSGKLAFETGGAGTQLMGRLXVMPAR  
REAPLCYRLNQCGLDRLPEVMAEASVDALADOTATOPALFTFEVALTAALRS  
MGVPELVAGHSIGELVAACVAGVFEPIVLAARGLMOLPGGAAVSLAADEA  
DVAANAAPHASVSLAANGPDQVVIAGACQPHALIAAANAAGARTKALHVSNAHS  
PLMAPLEAFGRVAVESVYRPSSTIVASNLISGAGDEVSFGVWRHAEEVFRPD  
VKALHAAGACFVVEGPKSTLGLVAPCLDPADPALAASSRGRDEPVALTEALGLM  
AVGSLVAGSRFPGSGRRVPLPTPMQRERYMIDTKAADARDBRAFGAHGEBVKG  
GAVGGDRSRALDHPPEPSGRREKVEAADRPRLIEDPGVLDRLVATREBRAG  
LGEVRIANDAGSLSPNDVALYVPPDLGCKNPMLLIGCEAGGAIIVANGEDVRIQ  
VGQVILASGACFATHTTTSALVYPPDLASATEAAMVATITMYALDGIARQV  
GERVLIHAATGVGLAAVQMAQHVGAHVHATAGPEKRAVLESLGVRVYSDSDRIV  
ADVATMGEGCEVDVNLISGELIDKSNLRSHRGVEGLCKCVAADQALGRLPRL  
NLNSELVDLRGMLERPARVALFEEELGSIAGVFPPEIATPLPVAADARSMAQ  
AOKGLVLTIVGDEVOIRIPTHAAGSPGEGDRLDRADGAPAAARAALEAFRL



[illegible]

Oy	181	ATGTCGCCCCCGCCAGCCGAGCGGAGCGCTCCGTTTCCCTCTACAGACATCCAAATATCC	240
Db	6440	ATGTGCCCCCGCCAGCCGAGCGGAGCGCTCCGTTTCCCTCTACAGACATCCAAAGATCC	6499
Oy	241	TACTGGCTGGGGCCGAGACAGAGAGCTTTATCGGTCCCGAGCGGATCCAGCTTATCGGAA	300
Db	6500	TACTGGCTGGGTCCGAGACAGAGAGCTTTAAGGTCCCGAGCGGATCCAGCTTATCGGAA	6555
Oy	301	TACGACTGTACGGAATCTGACGTGCCGAGGCTGAGCCGCGCTTTTCGAAAGTCGTCCGC	360
Db	6560	TACGACTGTACGGAATCTGACGTGGCCGAGGCTGAGCCGCGCTTTTCGAAAGTCGTCCGC	6611
Oy	361	CGGACGACATGCTTCCGGGGCCCAACAGCGTGGCCCGACATGATCAGATGTATCGAGCTTAA	420
Db	6620	CGGACGACATGCTTCCGGGGCCCAACAGCTGGCCCGACATGATCAGATGTATCGAGCTTAA	6677
Oy	421	GTGACGCGCCGACATCGAGATCATCGATCTGCGCGGGCTTCGACCGGAGCACAGGGAAAGCG	480
Db	6680	GTGACGCGCCGACATCGAGATCATCGATCTGCGCGGGCTTCGACCGGAGCACAGGGAAAGCG	6733
Oy	481	AGGCTCGGTGCTGTTCGAGATGCGGATGTCGACCGGCATCTAAGACACGAGGGCCCTCCG	540
Db	6740	AGGCTCGGTATCGTTTCGAGATGCGGATGTCGACCGGCATCTAAGACACGAGGCCCTCCG	6799
Oy	541	CTCTATCAGTGTGTCGCGCTTCGAGACGAGCGGCAAAACCGCTCGTCCGTCTAGTATC	600
Db	6800	CTCTATCAGTGTGTGCGGTTTCGCTGAGACGAGCAAAACCGCTCGTCCGTCTAGTATC	6855
Oy	601	GATCTCATTTAAGCTTGAACCTAGGACGCTGTCTCATATCTTCAAGACCTGGCTCAGCTTC	660
Db	6860	GATCTCATTTAAGTGTGACCTTAGGACGCTGTCTCATATCTTCAAGATTTGGCTCAGCTTC	6911
Oy	661	TACGAAATGCCGAGACCTCTCCCTGTCCGAGGCTCTGTACCGGATATATGATCTAC	720
Db	6920	TACGAAATGCCGAGACCTCTCTCCCTGTCTGTGAGGCTCTGTACCGGATATATGATCTAC	6977
Oy	721	GCGCTGAGTCTTCGCAAGAGTCTTGAGGCGCATCAACATCGATGATTTACTGAAAGCGG	780
Db	6980	GCGCTGAGTCTTCGCAAGAGTCTGAGGCGCATCAACATCGATGATTTACTGAAAGCGG	7033
Oy	781	CGCATTCGCGGAGCTCCCACTTCGCGGACGTTTCGAGTGAAGGCGGATCCATCTTACCCTG	840
Db	7040	CGCGTCCGAGCTCCCACTTCGCGGATGTTTCGAGTGAAGGCGGATCCATCTTACCCTG	7099
Oy	841	AAGAGATATCCGCTTCCGCGACACGAGAGATGGCTGCGGACTCTCGTGGGTGCATTTG	900
Db	7100	AAGAGATATCCGTTTCGCGACACGAGAGATGGCTGCGGACTCTCGTGGGTGCATTTG	7155
Oy	901	AAAGCGGCTGTCCGGGAGCGCGGGCTGACCCCGACGCGGCTCATCTTGCTGCATTTTCC	960
Db	7160	AAAGCGGCTGTCCGGGAGCGCGGGCTGACCCCGACGCGGCTCATCTTGCTGCATTTTCC	7211
Oy	961	GAGGTGATTCGGGCGCTGAGAGCGCGAGCCCGCGTTTAAAGCTCAATAAAGCTCTTCAAC	1022
Db	7220	GAGGTGATTCGGGCGCTGAGAGCGCGAGCCCGCGTTTAAAGCTCAATAAAGCTCTTCAAC	7277
Oy	1021	CGGCTCCCCGCTCATCCGCGCGTGAACGATATCACCGGGAACTTACGCTCATGTGCTTC	1088
Db	7280	CGGCTCCCCGCTCATCCGCGCGTGAACGATATCACCGGGAACTTACGCTCATGTGCTTC	7333
Oy	1081	CTGGAATCTGACACCATCTCCGACAAAGCTTTCGAACAGCGGCTTAAAGGTATTTCAAGG	1144
Db	7340	CTGGAATCTGACACCATCTCCGACAAAGCTTTCGAACAGCGGCTTAAAGGTATTTCAAGG	7399
Oy	1141	CAGCTGTGGGAAAGCATGATATCATCGACAGTAAAGGTATTCAGATGCTCAGCGAGAGGCC	1200
Db	7400	CAGCTGTGGGAAAGCATGATATCATCGACAGTAAAGGTATTCAGATGCTCAGCGAGAGGCC	7455
Oy	1201	GCCCCGGTCTTGGGGATCCAAAGAGGCGATTTTCCCCGTGTGCTCAAGCGCGCTT	1266
Db	7460	GCCCCGGTCTTGGGGATCCAAAGAGGCGATTTTCCCCGTGTGCTCAAGCGCGCGCTC	7511

```
QY 1261 AACGAGAGTGGTGTGTCACCTCGTTGACAGAGCTGGAACTCCGGGTGTACACCAAC 1320
DB |||||
QY 7520 AACCGAAGAGTGTGTGTCACTCTGTCAGAGAGCTGGCACTCCGGTGTACACCAAC 7579
DB |||||
QY 1321 AGCGAGACTCTCAGCTGTGTGATCATCACTCTACAGACAGATGGGACCTGTCTC 1380
DB |||||
QY 7580 AGCGAGACTCTCAGCTGTGTGATCATCACTCTACAGACAGATGGGACCTGTCTC 7639
DB |||||
QY 1381 CTGCGTGGGACATCGTGCAGAGAGTGTCCCGCCGACCTCTTGTGACGACATGTCTGAA 1440
DB |||||
QY 7640 CTGCGTGGGACATCGTGCAGAGAGTGTCCCGCCGACCTCTTGTGACGACATGTCTGAA 7699
DB |||||
QY 1441 GGGTACGTGTCTTCTCCGGGGGCTCACTGAGGAAACATGGGGTGAAGAGTGGCTGT 1500
DB |||||
QY 7700 GGGTACGTGTCTTCTCCGGGGGCTCACTGAGGAAACATGGAGTGAAGATGGCTGT 7759
DB |||||
QY 1501 TCGCTTCCGCTGCCAGCTAGAGAGCGCGGAGCGCAAAAGCGCAACCGCTGTCTG 1560
DB |||||
QY 7760 TCGCTTCCGCTGCCAGCTAGAGAGCGCGGAGCGCAAAAGCGCAACCGCTGTCTG 7819
DB |||||
QY 1561 ACGGAGCATACGCTGCAGCGGCTGTTCGGGGCGGGGTGAGCAGCTGCCATGCAGCTC 1620
DB |||||
QY 7820 ACGGAGCATACGCTGCAGCGGCTGTTCGGGGCGGGGTGAGCAGCTGCCATGCAGCTC 7879
DB |||||
QY 1621 GCGGTGTGTGCGGCGCGCAAGACGCTCACGTACGAAAGACTTTCGCGCGGTTCGCGGCA 1680
DB |||||
QY 7880 GCGGTGTGTGCGGCGCGCAAGACGCTCACGTACGAAAGACTTTCGCGCGGTTCGCGGCA 7939
DB |||||
QY 1681 CTTGGCGCGGCTGCGCGAGAGAGGGGAGACGCCCGCAACATTTGTCGCGGTGTGATG 1740
DB |||||
QY 7940 CTTGGCGCGGCTGCGCGAGAGAGGGGAGACGCCCGCAACATTTGTCGCGGTGTGATG 7999
DB |||||
QY 1741 GAGAAAGCTGGGAGAGAGGTTGTTCGGGTTCCTGCGGTCTGAGTCAAGCGCGCTAC 1800
DB |||||
QY 8000 GAGAAAGCTGGGAGAGAGGTTGTTCGGGTTCCTGCGGTCTGAGTCAAGCGCGCTAC 8059
DB |||||
QY 1801 GAGCGGATGATGCGGACCTTACCGGCGAGCGGATCACTACCTCTCATATGTTGAG 1860
DB |||||
QY 8060 GAGCGGATGATGCGGACCTTACCGGCGAGCGGATCACTACCTCTCATATGTTGAG 8119
DB |||||
QY 1861 GTTAAAGCTGTGTGACGAGCGCATGTGTGATGAGCAAACTGTCACTGCGCGGGATC 1920
DB |||||
QY 8120 GTTAAAGCTGTGTGACGAGCGCATGTGTGATGAGCAAACTGTCACTGCGCGGGATC 8179
DB |||||
QY 1921 CAGCGGCTGTGTGACGAGCGCGCGGTGTGAAAGCGAGCGGACCAAGCTTCGATGATG 1980
DB |||||
QY 8180 CAGCGGCTGTGTGACGAGCGCGCGGTGTGAAAGCGAGCGGACCAAGCTTCGATGATG 8239
DB |||||
QY 1981 CCCATTGACACCTTCGATCTGCGGTATGTCTATACCTCGGGATTCACAGGGTGTG 2040
DB |||||
QY 8240 CCCATTGACACCTTCGATCTGCGGTATGTCTATACCTCGGGATTCACAGGGTGTG 8299
DB |||||
QY 2041 CCCAAGGGGGTATGATCATCATCGGGGTGCGGTCAACCACTCTGACATCAACGAG 2100
DB |||||
QY 8300 CCCAAGGGGGTATGATCATCATCGGGGTGCGGTCAACCACTCTGACATCAACGAG 8359
DB |||||
QY 2101 CGCTTGAATAAGGCGCGGAGACAGAGGTGTGTGCGCTCTCTCGGTGAGCTTCATCTC 2160
DB |||||
QY 8360 CGCTTGAATAAGGCGCGGAGACAGAGGTGTGTGCGCTCTCTCGGTGAGCTTCATCTC 8419
DB |||||
QY 2161 TCGGATATGATGTGTTCGGGATCTCGGCGCGGGGCGGTACGATCGTGTGCGGAGCG 2220
DB |||||
QY 8420 TCGGATATGATGTGTTCGGGATCTCGGCGCGGGGCGGTACGATCGTGTGCGGAGCG 8479
DB |||||
QY 2221 TCCAAAGTTCGCGATCGGCGCATTTGGGAGAGTTGATGAAAGAGAGAGTGCAGTGTG 2280
DB |||||
QY 8480 TCCAAAGTTCGCGATCGGCGCATTTGGGAGAGTTGATGAAAGAGAGTGCAGTGTG 8539
DB |||||
QY 2281 TGGAACTCGGTGCGCGCGCTGATGCGGATGCTGTGAGCATTTTGAAGGTGCGCCGAT 2340
DB |||||
QY 8540 TGGAACTCGGTGCGCGCGCTGATGCGGATGCTGTGAGCATTTTCCAGAGGTGCGCCGAT 8599
DB |||||
QY 2341 TGGCTGTAGAGTCTGTGGGGCTTTCGCTGTGAGGCGGAGCTGATCCCGGTGGGCTGTG 2400
DB |||||
DB 8600 TCGCTGTAGAGTCTGTGGGCTTTCGCTGTGAGGCGGAGTGGATCCCGGTGGGCTGTG 8659
QY 2401 CTTGCGAGCTTCAGAGCCATCAAGGCCCGGCGGTGTGTGATCAAGCTTGGGCGGAGCAAC 2460
DB |||||
QY 8660 CTTGCGAGCTTCAGAGCCATCAAGGCCCGGCGGTGTGTGATCAAGCTTGGGCGGAGCAAC 8719
DB |||||
QY 2461 GAAAGTGTGATTTGTGATCAAGGGTATCCCGGTGAGGAAAGTTCGATCTGTGTGGGCGAGC 2520
DB |||||
QY 8720 GAAAGTGTGATTTGTGATCAAGGGTATCCCGGTGAGGAAAGTTCGATCTGTGTGGGCGAGC 8779
DB |||||
QY 2521 ATCCCTACGAGCGGCTGCGGAGCAACAGAGAGTTCACAGTGTGATGAGGCGCTCGAA 2580
DB |||||
QY 8780 ATCCCTACGAGCGGCTGCGGAGCAACAGAGAGTTCACAGTGTGATGAGGCGCTCGAA 8839
DB |||||
QY 2581 CCGGCGCGGCTGTGGGATTCGGGGCACTTACATTTGGCGGGGTGAGCTGTGCACTGGGC 2640
DB |||||
QY 8840 CCGGCGCGGCTGTGGGATTCGGGGCACTTACATTTGGCGGGGTGAGCTGTGCACTGGGC 8899
DB |||||
QY 2641 TACTGGCGGATGAAAGAGAGAGCGGAGAGCTTCTCGTGTGACCCCGAGACCGGGAG 2700
DB |||||
QY 8900 TACTGGCGGATGAAAGAGAGAGCGGAGAGCTTCTCGTGTGACCCCGAGACCGGGAG 8959
DB |||||
QY 2701 GCGCTCTACAGAGCGGCGATCTGGGCGGCTACCTGCGGATGGAATGAAATCATGAGTTATG 2760
DB |||||
QY 8960 GCGCTCTACAGAGCGGCGATCTGGGCGGCTACCTGCGGATGGAATGAAATCATGAGTTATG 9019
DB |||||
QY 2761 GGGCGTGTGAGACCAACCAATCAAGCTTTCGGGATTCGCGGTTGAGCTGTGGGAAATCAG 2820
DB |||||
QY 9020 GGGCGGAGAGCAACCAATCAAGCTTTCGGGATTCGCGGTTGAGCTGTGGGAAATCAG 9079
DB |||||
QY 2821 GAAAGCTCAAGTGTGATTCGGAACCTACGCGAGCGGGGTGATTTGGCCGTGGGAGACAG 2880
DB |||||
QY 9080 GAAAGCTCAAGTGTGATTCGGAACCTACGCGAGCGGGGTGATTTGGCCGTGGGAGACAG 9139
DB |||||
QY 2881 GCGGCGAACAAGCTCTCTTACGCTATGTGTGTCGCGAGGAGCAACAGAGACGCGCTGC 2940
DB |||||
QY 9140 GCGGCGAACAAGCTCTCTTACGCTATGTGTGTCGCGAGGAGCAACAGAGACGCGCTGC 9199
DB |||||
QY 2941 GAGCAGAGCGGAGCTTCAAGACCGAGGATTCAGCGGAGACACAGCGCGCGAGACG 3000
DB |||||
QY 9200 GAGCAGAGCGGAGCTTCAAGACCGAGGATTCAGCGGAGACACAGCGCGCGAGACG 9259
DB |||||
QY 3001 GAGCGCTTGTGAGAGGAGGAGAGGAGTGTGAGTCAAGCTTCGCTGCAAGAGCTTCGAGG 3060
DB |||||
QY 9260 GAGCGATTGTGAGAGGAGGAGGAGTGTGAGTCAAGCTTCGCTGCAAGAGCTTCGAGG 9319
DB |||||
QY 3061 GACTGTGACGAGAAAGCCGCTGTGATGTGACCGGGCAGATTCGCGGAGGCGGGGCTGTG 3120
DB |||||
QY 9320 GACTGTGACGAGAAAGCCGCTGTGATGTGACCGGGGCTGTGACCGGGGAGGCGGGGCTGTG 9379
DB |||||
QY 3121 GAGCTGTACGCGCGGTGTGAGCTTCGGAACGTTCTTTGAGCGCCGATTCGTTGT 3180
DB |||||
QY 9380 GAGCTGTACGCGCGGTGTGAGCTTCGGAACGTTCTTTGAGCGCCGATTCGTTGT 9439
DB |||||
QY 3181 GAGTTGTGATTCATCTGAGGCTGTGAGAGCGGTGAGAGCGGAGCGGAGCGGAGCGCTTCC 3240
DB |||||
QY 9440 GAGTTGTGATTCATCTGAGGCTGTGAGAGCGGTGAGAGCGGAGCGGAGCGGAGCGCTTCC 9499
DB |||||
QY 3241 AAATTCGTTATTCATCTGCGCGGAGCAAGTACCGGATGCAAACTCATGCGTATGTCAA 3300
DB |||||
QY 9500 AAATTCGTTATTCATCTGCGCGGAGCAAGTACCGGATGCAAACTCATGCGTATGTCAA 9559
DB |||||
QY 3301 TCCGCGCGGATGTGAGGCGGTGTGAGAGGAGCTTCTATTTATTTACACCGCTTGTGAGCAAG 3360
DB |||||
QY 9560 TCCGCGCGGATGTGAGGCGGTGTGAGAGGAGCTTCTATTTATTTACACCGCTTGTGAGCAAG 9619
DB |||||
QY 3361 TTTGCTGAAGCTTCCGATCAAGGAGTTCAGAGCGGAGGAGCGGAGCGGAGCGGAGCGGAG 3420
DB |||||
QY 9620 TTTGCTGAAGCTTCCGATCAAGGAGTTCAGAGCGGAGGAGCGGAGCGGAGCGGAGCGGAG 9679
DB |||||
QY 3421 GTTGTGTGAAAGCGGCTTCAACTCTGTGTTGTTGAGGAGGAGTTCAGGCGCATGAGTGTG 3480
DB |||||
```

Db 9680 GTGTTGATGAAAGCGGCTTGCGCTCTGTTGTTGAGGAGATGATGATCCATGAGTCG 9739  
Qy 3481 CTGATGATGATGCTGCTGCGAGAAATTTTGCTGTGAGGCGGATATATAGCGAGCTC 3540  
Db 9740 CTGATGATGATGCTGCTGCGAGAAATTTTGCTGTGAGGCGGATATATAGCGAGCTC 9799  
Qy 3541 CTGATGAGCAAGGCGCTTCTGCAACATGCGGCTGTCTCGGTGGGCAATTCATTTT 3600  
Db 9800 CTGATGAGCAAGGCGCTTCTGCAACATGCGGCTGTCTCGGTGGGCAATTCATTTT 9859  
Qy 3601 GAAAGGTTGCGGCGGCTTCTGCAACATGCGGCTGTCTCGGTGGGCAATTCATTTT 3660  
Db 9860 GAAAGGTTGCGGCGGCTTCTGCAACATGCGGCTGTCTCGGTGGGCAATTCATTTT 9919  
Qy 3661 GCGCGGCGGATAGACCGCGGAGTTCCAGGCTGTGAGCTGAGTCCAGATTTCTCAACG 3720  
Db 9920 GCGCGGCGGATAGACCGCGGAGTTCCAGGCTGTGAGCTGAGTCCAGATTTCTCAACG 9979  
Qy 3721 AGGCGCGCAAGCGCGGCGGCTTCCGCGCGGAGAGCACTTCCGAGATATGCTT 3780  
Db 9980 AGGCGCGCAAGCGCGGCGGCTTCCGCGCGGAGAGCACTTCCGAGATATGCTT 10039  
Qy 3781 CGGCACTTCTGAGAGCAACAACTACCGAGTACATGAGTCCCTACAGTCTTGTGAGCTC 3840  
Db 10040 CGGCACTTCTGAGAGCAACAACTACCGAGTACATGAGTCCCTACAGTCTTGTGAGCTC 10099  
Qy 3841 GATGCGTTGCGGCTGACCTCCAAAGGAGTGCATGCTGTAAGGCTTCCGAGCGAG 3900  
Db 10100 GATGCGTTGCGGCTGACCTCCAAAGGAGTGCATGCTGTAAGGCTTCCGAGCGAG 10159  
Qy 3901 GATGCTGTCGCGCGGCTGACCTCCAAAGGAGTGCATGCTGTAAGGCTTCCGAGCGAG 3960  
Db 10160 GATGCTGTCGCGCGGCTGACCTCCAAAGGAGTGCATGCTGTAAGGAGATC 10219  
Qy 3961 CTCGTCGCGGCTGACCTCCAAAGGAGTGCATGCTGTAAGGCTTCCGAGCGAGCTTC 4020  
Db 10220 CTCGTCGCGGCTGACCTCCAAAGGAGTGCATGCTGTAAGGCTTCCGAGCGAGCTTC 10279  
Qy 4021 GTCGATCTTGTGCGACATGATTCATGCTGTCATGAGAGGCTTTCAGAGAGG 4080  
Db 10280 GTCGATCTTGTGCGACATGATTCATGCTGTCATGAGAGGCTTTCAGAGAGG 10339  
Qy 4081 CTGATGAGGAGATGCGCATCACGAGTGTTCAGTACCGGAACTTCGCTGCTGCG 4140  
Db 10340 CTGATGAGGAGATGCGCATCACGAGTGTTCAGTACCGGAACTTCGCTGCTGCG 10399  
Qy 4141 TCCGTTTTCGCGCGGAGATCTGAGAGATCTAGATGAGCGCGGAGATCAGAGCGAGTG 4200  
Db 10400 TCCGTTTTCGCGCGGAGATCTGAGAGATCTAGATGAGCGCGGAGATCAGAGCGAGTG 10459  
Qy 4201 GAGGTTGCGCGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4233  
Db 10460 GAGGTTGCGCGGAG 10492

RESULT 9  
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001  
DEFINITION Sequence 2 from patent US 6303342.  
ACCESSION AR172664  
VERSION AR172664.1 GI:17912155  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 71989)  
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.  
TITLE Recombinant methods and materials for producing epothenones C and D  
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;  
FEATURES location/Qualifiers  
source 1..71989  
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN  
Query Match 97.2%; Score 4114.6; DB 6; Length 71989;  
Best Local Similarity 98.3%; Pred. No. 0;  
Matches 4159; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
Qy 1 ATGACGATCAATGACTTCTGAAAGAGCTGAGACAGAGGATATCAAGCTGCGCGCAT 60.  
Db 6260 ATGACGATCAATGACTTCTGAAAGAGCTGAGACAGAGGATATCAAGCTGCGCGCAT 6319  
Qy 61 GGGAGAGGCTTCCAGATACAGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db 6320 GGGAGAGGCTTCCAGATACAGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6379  
Qy 121 ATTCGAGCACAAAAGAGAGATCTGAGAGAGTGTCTGCGTGAAGATCTCCGAGAAATCC 180  
Db 6380 ATTCGAGCACAAAAGAGAGATCTGAGAGAGTGTCTGCGTGAAGATCTCCGAGAGATCC 6439  
Qy 181 ATCGTCCCGCGCGGCGGAG 240  
Db 6440 ATCGTCCCGCGCGGCGGAG 6499  
Qy 241 TACTGCGTGGCGGAG 300  
Db 6500 TACTGCGTGGCGGAG 6559  
Qy 301 TACGACTGTACGAGTCTGAGAGTGCAGAGCTGAGAGCGGCGCTTCCGAAAGTGTGCGG 360  
Db 6560 TACGACTGTACGAGTCTGAGAGTGCAGAGCTGAGAGCGGCGCTTCCGAAAGTGTGCGG 6619  
Qy 361 CGGAGAGAGATGCTTCCGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Db 6620 CGGAGAGAGATGCTTCCGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6679  
Qy 421 GTCGACCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
Db 6680 GTCGACCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6739  
Qy 481 AGGCTGCTGCTGCGAGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Db 6740 AGGCTGCTGCTGCGAGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6799  
Qy 541 CTCTATCAGCTGCTGCGGCTGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
Db 6800 CTCTATCAGCTGCTGCGGCTGCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6859  
Qy 601 GATCTCATTAACGTTGACCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 6860 GATCTCATTAACGTTGACCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6919  
Qy 661 TACGAGATCCGAG 720  
Db 6920 TACGAGATCCGAG 7279  
Qy 721 GCGCTGAGTCTGCAAG 780  
Db 6980 GCGCTGAGTCTGCAAG 7039  
Qy 781 GCGTGCAG 840  
Db 7040 GCGTGCAG 7099  
Qy 841 AAGGAGATCCGCTTCCGAG 900  
Db 7100 AAGGAGATCCGCTTCCGAG 7159  
Qy 901 AAGGCGGCTGCGGAG 960  
Db 7160 AAGGCGGCTGCGGAG 7219  
Qy 961 GAGGTGATCGGCGCTGAG 1020  
Db 7220 GAGGTGATCGGCGCTGAG 7279

QY 1021 CCGCTCCCGCTGCATCCGCGCTGAACGATATCAACGGGAGCTTTCACGTCGATGCTCTC 1080  
DB CCGCTCCCGCTGCATCCGCGCTGAACGATATCAACGGGAGCTTTCACGTCGATGCTCTC 7339  
QY 1081 CTGACATGACACCACTGCGCAACAAGACTTCGAACAGCGGCTTAAGGCTATTCAGAG 1140  
DB CTGACATGACACCACTGCGCAACAAGACTTCGAACAGCGGCTTAAGGCTATTCAGAG 7399  
QY 1141 CAGCTGTGGAAACGATGATCACTGCGAGTAAAGCGTATCGAGTTCAGCGAGAGGC 1200  
DB CAGCTGTGGAAACGATGATCACTGCGAGTAAAGCGTATCGAGTTCAGCGAGAGGC 7459  
QY 1201 GCGCGGCTCTGGGATCCAAAGAGCGGATGTTCCCGTGGTGTCAAGAGCGGCTT 1260  
DB GCGCGGCTCTGGGATCCAAAGAGCGGATGTTCCCGTGGTGTCAAGAGCGGCTT 7519  
QY 1261 AACCAAGATGCTGTGCTCACTGCTGCAAGAGGCTGGAATCCGCTGTACACAGC 1320  
DB AACCAAGATGCTGTGCTCACTGCTGCAAGAGGCTGGAATCCGCTGTACACAGC 7579  
QY 1321 ACGGACACTCTCAGCTGCTGTGATCATCAGCTTACAGACAGATGAGGAGCCTGCTC 1380  
DB ACGGACACTCTCAGCTGCTGTGATCATCAGCTTACAGACAGATGAGGAGCCTGCTC 7639  
QY 1381 CTGCGTGGGACATGCTGCAAGAGTGTCCCGCCGACCTTCTGAGACGATGCTCGAA 1440  
DB CTGCGTGGGACATGCTGCAAGAGTGTCCCGCCGACCTTCTGAGACGATGCTCGAA 7699  
QY 1441 GCGTACGTGCTTTTCTCCGCGGCTCATCTGAGAAACATGAGGAGTGAACAGGCTGCT 1500  
DB GCGTACGTGCTTTTCTCCGCGGCTCATCTGAGAAACATGAGGAGTGAACAGGCTGCT 7759  
QY 1501 TCGCTTCGCGCTGCCAGCTAAGAGCGGCGAGCAACGCGACCAACGCGCTGCTG 1560  
DB TCGCTTCGCGCTGCCAGCTAAGAGCGGCGAGCAACGCGACCAACGCGCTGCTG 7819  
QY 1561 ACGGACATACGCTGCAAGGCTGTTCCGCGGCGGCTGAGACGCTGCCATGACGCTC 1620  
DB ACGGACATACGCTGCAAGGCTGTTCCGCGGCGGCTGAGACGCTGCCATGACGCTC 7879  
QY 1621 GCGTGTGTGCTGCGCGCAAGAGCTCATGTAAGAAAGCTTTCGCGGCTTCCGCGCA 1680  
DB GCGTGTGTGCTGCGCGCAAGAGCTCATGTAAGAAAGCTTTCGCGGCTTCCGCGCA 7939  
QY 1681 CTGCGCGCGCTGCGAGAGAGGCGGAGCAAGCTTGTGCTGCGGCTGATG 1740  
DB CTGCGCGCGCTGCGAGAGAGGCGGAGCAAGCTTGTGCTGCGGCTGATG 7999  
QY 1740 GAGAAAGCTGAGAGAGGTTTCCGCGGCTTCCGCGGCTGATGAG 1800  
DB GAGAAAGCTGAGAGAGGTTTCCGCGGCTTCCGCGGCTGATGAG 8059  
QY 1801 GTGCGGATGATGCGCACTTACCGGCGAGAGCTATCACTACCTCTCATGATGATG 1860  
DB GTGCGGATGATGCGCACTTACCGGCGAGAGCTATCACTACCTCTCATGATGATG 8119  
QY 1861 GTAAGCTGTGCTGACGAGCATGCTGATGAGCAACTGTATGCTGCTGATGATG 1920  
DB GTAAGCTGTGCTGACGAGCATGCTGATGAGCAACTGTATGCTGCTGATGATG 8179  
QY 1921 CAGCGGCTGTGAGAGAGCGCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
DB CAGCGGCTGTGAGAGAGCGCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8239  
QY 1981 CCGATTCAGACACTTCGATCTCGGATGCTATCACTTCAGGATTCAGAGGTTG 2040  
DB CCGATTCAGACACTTCGATCTCGGATGCTATCACTTCAGGATTCAGAGGTTG 8299  
QY 2041 CCGAAGGAGTATGATCATCGAGGCTGCTCAACACCATCTTGAACATCAACGAG 2100  
DB CCGAAGGAGTATGATCATCGAGGCTGCTCAACACCATCTTGAACATCAACGAG 8359

QY 2101 CCGTGAATAGAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
DB CCGTGAATAGAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8419  
QY 2161 TCGCTGTATGATGCTGAGGATCTGAGCGGCGGAGTACGATGCTGCTGAGAGAGAG 2220  
DB TCGCTGTATGATGCTGAGGATCTGAGCGGCGGAGTACGATGCTGCTGAGAGAGAG 8479  
QY 2221 TCCAAAGCTGCGGATCCGCGGCTTGGGAGAGAGTTCGAACAGAGAGAGAGAGAGAG 2280  
DB TCCAAAGCTGCGGATCCGCGGCTTGGGAGAGAGTTCGAACAGAGAGAGAGAGAGAG 8539  
QY 2281 TGAAGCTGCGGAGTCCGCGGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
DB TGAAGCTGCGGAGTCCGCGGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8599  
QY 2341 TCGCTGTAGGCTCTGAGCGGCTTTCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400  
DB TCGCTGTAGGCTCTGAGCGGCTTTCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 8659  
QY 2401 CCGTGGAGCTTCAGGACATCAGGCGCGGCTGCTGATGATCAGCTGCGGCGGAGAG 2460  
DB CCGTGGAGCTTCAGGACATCAGGCGCGGCTGCTGATGATCAGCTGCGGCGGAGAG 8719  
QY 2461 GAAGGCTGATCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520  
DB GAAGGCTGATCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8779  
QY 2521 ATCCCTTACGCGCGCTGCGCTGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580  
DB ATCCCTTACGCGCGCTGCGCTGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 8839  
QY 2581 CCGCGCGGCTGCTGAGGTTCCGCGGAGACTTCACTTATGAGAGAGAGAGAGAGAGAG 2640  
DB CCGCGCGGCTGCTGAGGTTCCGCGGAGACTTCACTTATGAGAGAGAGAGAGAGAGAG 8899  
QY 2641 TACTGCGCGATGAG 2700  
DB TACTGCGCGATGAG 8959  
QY 2701 CCGCTTCAAGAGCGGAGATCTGAGCGGCTTACCTGCGGAGAGAGAGAGAGAGAGAG 2760  
DB CCGCTTCAAGAGCGGAGATCTGAGCGGCTTACCTGCGGAGAGAGAGAGAGAGAGAG 9019  
QY 2761 GCGCGTGAAGACCAATCAAGCTTCCGCGGATACCGCGGTTGAGCTGCGGAGAGAG 2820  
DB GCGCGTGAAGACCAATCAAGCTTCCGCGGATACCGCGGTTGAGCTGCGGAGAGAG 9079  
QY 2821 GAAGCGCTCAAGTCCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880  
DB GAAGCGCTCAAGTCCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9139  
QY 2881 GCGGAGAAAGAGCTCTTCTAGCTTATGCTGCTGAGAGAGAGAGAGAGAGAGAGAG 2940  
DB GCGGAGAAAGAGCTCTTCTAGCTTATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 9199  
QY 2941 GAGCAGAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000  
DB GAGCAGAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9259  
QY 3001 GAGCGCTTGAAG 3060  
DB GAGCGCTTGAAG 9319  
QY 3061 GACCTGAG 3120  
DB GACCTGAG 9379  
QY 3121 GACGCTTGAAG 3180  
DB GACGCTTGAAG 9439  
QY 3181 GAGTTTGTGATCTCTGAGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240

Db 9440 GAAATTCGGCCCATCTGAGCTGCTGAGCAGCGGAGCCGACGCGCGCCCTTCC 9499  
 Qy 3241 AAATTCGTTATTCATGCGCGCGGCGAGCAGTACCCCGGCAAACTAGCCGTATGCAA 3300  
 Db 9500 AAATTCGTTATTCATGCGCGCGGCGAGCAGTACCCCGGCAAACTAGCCGTATGCAA 9559  
 Qy 3301 TCCGCGCCGATCGAGGCGGTGAGCAGAGGCTTATTAATTAACAACCCGTTGAGACCCGT 3360  
 Db 9560 TCCGCGCCGATCGAGGCGGTGAGCAGAGGCTTATTAATTAACAACCCGTTGAGACCCGT 9619  
 Qy 3361 TTGCTGAAGCTCTCCGATCAGCGGATGAGCGCGGAGCGCAGTTCCGCAAACTTCGAC 3420  
 Db 9620 TTGCTGAAGCTCTCCGATCAGCGGATGAGCGCGGAGCGCAGTTCCGCAAACTTCGAC 9679  
 Qy 3421 GTGTTCGATGAAGGCGGCTTCAACCTCTGTTCCGCGGAGAGTACGAGCCATGAGTCG 3480  
 Db 9680 GTGTTCGATGAAGGCGGCTTCAACCTCTGTTCCGCGGAGAGTACGAGCCATGAGTCG 9739  
 Qy 3481 CTGTATGAGATCGTCTGCGAGAAATTTTCCCTGCTGAGGCGGAGTATATGAGCGAGTC 3540  
 Db 9740 CTGTATGAGATCGTCTGCGAGAAATTTTCCCTGCTGAGGCGGAGTATATGAGCGAGTC 9799  
 Qy 3541 CTGTATGAGATCGTCTGCGAGAAATTTTCCCTGCTGAGGCGGAGTATATGAGCGAGTC 3600  
 Db 9800 CTGTATGAGATCGTCTGCGAGAAATTTTCCCTGCTGAGGCGGAGTATATGAGCGAGTC 9859  
 Qy 3601 GAAAGGTTGCGCGCGGCTTCTGAGACCTGAGACATTCGAGCGGATTAAGTCGAGTCG 3660  
 Db 9860 GAAAGGTTGCGCGCGGCTTCTGAGACCTGAGACATTCGAGCGGATTAAGTCGAGTCG 9919  
 Qy 3661 GCGCGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 3720  
 Db 9920 GCGCGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 9979  
 Qy 3721 AGCGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 3780  
 Db 9980 AGCGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 10039  
 Qy 3781 CGCGACCTTCTGAGAGCAAACTACCCGAGTACATGAGTCTTCAAGTCTTCTGAGTC 3840  
 Db 10040 CGCGACCTTCTGAGAGCAAACTACCCGAGTACATGAGTCTTCAAGTCTTCTGAGTC 10099  
 Qy 3841 GATCGGTTGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 3900  
 Db 10100 GATCGGTTGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 10159  
 Qy 3901 GATACTTCTGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 3960  
 Db 10160 GATACTTCTGCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 10219  
 Qy 3961 CTGCTCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 4020  
 Db 10220 CTGCTCGCGGATGAGACCCGCGGAGTTCAGAGTCTGTAAGTCGAGTTCCTCAACG 10279  
 Qy 4021 GTGATCTTGTGAGACATTCATCGTTCGATGAGAGGCTTGTGAGAGAGG 4080  
 Db 10280 GTGATCTTGTGAGACATTCATCGTTCGATGAGAGGCTTGTGAGAGAGG 10339  
 Qy 4081 CTGATGAGAGATTCGATCAACCGAGTTCAGTACCCGAACTTCGAGTCGTCGCG 4140  
 Db 10340 CTGATGAGAGATTCGATCAACCGAGTTCAGTACCCGAACTTCGAGTCGTCGCG 10399  
 Qy 4141 TCCGTTTGGCGCGAGATTCAGAGATTCATGAGCGCGCGGAACTTCAGAGATTCGAGTCG 4200  
 Db 10400 TCCGTTTGGCGCGAGATTCAGAGATTCATGAGCGCGCGGAACTTCAGAGATTCGAGTCG 10459  
 Qy 4201 GAGGTTGCGCGCGAGAGCGAGAGCTAGCTAA 4233  
 Db 10460 GAGGTTGCGCGCGAGAGCGAGAGCTAGCTAA 10492

RESULT 10

AF188287  
 LOCUS 42603 bp DNA linear BCT 12-JAN-2000  
 DEFINITION Stigmatella aurantiaca myxothiazol gene cluster, complete sequence.  
 ACCESSION AP188287  
 VERSION AP188287.1 GI:6635393  
 KEYWORDS  
 ORGANISM  
 SOURCE  
 Stigmatella aurantiaca  
 Stigmatella aurantiaca  
 Bacteria: Proteobacteria: Deltaproteobacteria: Myxococcales;  
 Cyctobacteriineae; Cyctobacteraceae; Stigmatella.  
 REFERENCE  
 1 (bases 1 to 42603)  
 Sliakowski, B., Schaiter, H.U., Ehret, H., Kunze, B., Weinig, S.,  
 Nordstedt, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S. and  
 Miller, R.  
 New lessons for combinatorial biosynthesis from myxobacteria. The  
 myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca  
 DM4/3-1  
 JOURNAL  
 J. Biol. Chem. 274 (52), 37391-37399 (1999)  
 MEDLINE  
 20069734  
 PUBMED  
 10601310  
 REFERENCE  
 2 (bases 1 to 42603)  
 Sliakowski, B., Schaiter, H.U., Ehret, H., Kunze, B., Weinig, S.,  
 Nordstedt, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S. and  
 Miller, R.  
 Direct Submission  
 Submitted (20-SEP-1999) NBI/NC, GBF, Mascheroder Weg 1,  
 Braunschweig 38124, Germany  
 FEATURES  
 source  
 Location/Qualifiers  
 1..42603  
 /organism="Stigmatella aurantiaca"  
 /mol\_type="genomic DNA"  
 /strain="DM4/3-1"  
 /db\_xref="taxon:41"  
 1..834  
 /gene="mtaA"  
 1..834  
 /gene="mtaA"  
 /note="Pant transferase"  
 /codon\_start=1  
 /transl\_table=1  
 /product="mtaA"  
 /protein\_id="AAPI9809.1"  
 /db\_xref="GI:6635394"  
 /translation="MPTSPDALPLKLPDPDVHVIPEPERITGLESYRALDPC  
 ERDQKQFPERHRIQVLSHVALVRLISVAPAPAMGFSANQGRPEIRGEKPM  
 LRLNLSHTDMALCANVARDVDGADVDTERRGRTVIAISFPAVASELRALPVG  
 QREFFPDYWLKEAYIVAKGMGLSLPDQAFVSGQLSTRISFDPLVDEPSQWQV  
 RPRSORHAAALAVRSEAPLVRFQRTVPLDDAPAEYLSRRRIQPLRLRMVGVG  
 G"  
 1388..13399  
 /gene="mtaB"  
 /note="PKS"  
 1388..13399  
 /gene="mtaB"  
 /note="PKS; MtaB loading domain and module 1: polyketide  
 synthase domains: ACP, KS, AT, AT, DH, ER, KR, ACP, MtaB  
 module 2: polyketide synthase domains: KS, AT, DH, KR,  
 ACP"  
 /codon\_start=1  
 /transl\_table=1  
 /product="mtaB"  
 /protein\_id="AAPI9810.1"  
 /db\_xref="GI:6635395"  
 /translation="MSTHEVGVGVRSTGQLRDMVNOQVAVSONGLBAHTVIRTPPS  
 LGVGSVPATRLIAELARLPKPLSPVLVWHPITIDKARLISGAEPRAEMHARGEPA  
 AVDGAELIALVGLACRPKAPDAKAFWRVQEDRCVTSVSNRMDSLIYDPAVGP  
 GKTIVTRGALVDSVDRDPVFGISPREACQLDPOQRLMELSEALELDGIAFVIR  
 GSRTGVGTGMVMDVADLQTIHNAAPLTHCTGAGGI IANRISYALGEGSLAIDT  
 ACSSSLVAIHEACRLSRGSSSLAAGVQVLSPTRFMAVSAFALAPGRCTFPA  
 RASGYVRGEGGIVLKLPSRALADGAIYICLRGSVYVNDGASNGLTANPRKQELT  
 LRDAVYVAGMDPAQVHVELHGTGTPLGDETEGALGAVIGAAEARPRLHVGSKTN  
 IGHLEAAAGIASIVIKVALSIRHOOLPASLHPEQNPRIIPDDILRLVQTLAPPAK  
 DPVAVGISAFPGFTNCHVVVEHRGGSRTAAVPSVTPHQSFRHPPTAFLFAGAGSA

WMGMBEELIAGAEVFAREPCDAAAPRLGSMGLVBAALAPVRSGRPGGRODLIVL  
 AFQISLAALMRWGIERSPATVTCYSGEAAAAYAGLITLIEEPALATHTIOUKVAC  
 GVMALVLSGAEVARIAPYGRISISENGPRSTGJGEAAALQDFMTATTAAGT  
 ALVLEIDVYAHPOVEALKPLFQSPVLPSPGPRVMTVSAGFVPGCTGAGAWA  
 SNREPRPATAMRLTBEGYDVLDVNPPIHESIVLENRAGTASAEFISMGRG  
 PVEGHRALITOLSOAGGFWERVVDVPGACSSAPVRETELPAVARTFOALREODR  
 IAEOLRDAASLTDLCYTAGARGSHHEHRLVLAETRALBEGGLAARPGEIRNPS  
 QCMVRGEPRLGPVPSGGSFPLSSVLFEEBPSFRAPRERSEIRPAOQASLIE  
 ALHAERESGHPRETEIEPDLFPAFOALLAWESWGIRPAVAGHVEISAHVAGV  
 ILSLEALSLVSRSLMGRAGAGQMAVTEASPARIEPLGCGTDLAAANGEMBY  
 VVSGALAAALASLDELITQGLRVRVLNVYAFHSQMAPYLGDLPLVLELKPPEPL  
 PLYTHGRSLASDEGTDHRAVNERVLPAPEAMRABDGHAFVESIPPLVREL  
 LSRIIDASPASGASALFSTRBRAPRAALIGLITLTLGHRSMKGVPEAGTIVEL  
 BSYMNORQCWFDAAPQALAEVPHLGRVYAPPGQYFEATFRSDRLQILSDIRLY  
 RLVVAGASGLQMLVAAQOLWGEAPCALQDVLITPVLGDTSEORVLSATGAG  
 RILSWAGQGGEGTSSHMI BHATGRLERGAEETEGREVLSEARCRFLASAEFE  
 GIARQGLIGRQFRLVLELMQDREAFGLRPEREDDGAEDYRFGGLDSCFOYGA  
 VAGPQIEGLAFVPIGVGRFALRGEARPPRFGHLVLRGSGKANEASVDLILDESG  
 LVAAVSGLVIRRAPREVMRLRRDVTAPPLVLTETMQPOMLSGASOAGPRLMLADRG  
 GLAARLBOLSGRGRVLYLAABENAOGLGRIEMADAGLACAYVDRLGIDRESS  
 DAVQKVKSVQOLVTRRAEALRLVLTGTQVTSQDGLIEBQPLMGLDNVLALE  
 HPELCQRIDLPSPGAPDAQALAEALAGREDRVAFEGRLVARTLRTLPVAVR  
 ALPLEBDATVYLIVGIGIGIELARWVGARHLVLMGRSAPSAAQOTTLALDAG  
 ATISVALGNVAREALAVLGDIRAMPPLRGVMAAGVLDGLVROSEERLAVLA  
 PRTIOGMNTHRETVGDRDLPEFLFSSVSFQTSQAGYASANAFLDALHRRROGL  
 TALSTMGMDVGMMAARLPBAGKSSFRGILASPAGLAEIRLLSSEAQYVVA  
 GASBAKCDPAPFFDALSRRKPEASRPVUGALVAGAPRELRALLSLLOVATV  
 LCLPSPAPIDPRPLOELGLDLMALQFNNSLAASIGRSLPATLLFDHETLEKLEFL  
 ADLTPASEASPAKSRPVKLPARDAEGAEPLAVVGMACRLPGSVASPEEFRLAMG  
 VNAVEVPKDRMSLDTWYDPDETPCKMYSRHGDFEVDPPDAPFGISAEASAMG  
 POORLIELGWEALSAGIADOLKSGPFGVGLICSDYALLETINSPPRTINYSG  
 SGGVSVAGRIAYALGLEGPTLADTPACSSIVNAHLACOSLRNGECILAYGNSL  
 LLSPRKSVTFSTKALSPDACKAPDSSANGVYRGAGVLYKRLSEBLADGDTLS  
 VWRGAVNODGRSNGITAPRAQOVERVIESALRQGVSPLDVSVYAEHGTSLDPI  
 EVOALAAVNGKDRPSAPLIGSVKTVNGHLEGAAGVNLKAIITALHRELRPSLHF  
 KTPSPHIMDELPVEVISHRPMRWVEGKRILIGVSGFSGTNHVLIEEAPENGR  
 SKAPSIPEERBELVLSARAPALOEAAVEFSAOLGRAGDAAMLHDVCSKGRSH  
 HEHRLAVVARSSTEAALAEATLGLAHPDNOGYALBGAIPRVVLTPLFGOGGCMPT  
 GHALLEBEPFRSALOADEVILKLAGNSITIELORETFRPQMAISOPARALDSAL  
 LALLRSWGIEPDAVAGSSIGEVAAAYAAVLTLEEMRTVARSRLIHAVAKGAMAL  
 VLSSEBATEVAAAYAOQLSVAAINSQNTSLLSGEAGMDELGALTAGVSCRIGL  
 DYASHPAVEPYIDELKQLOGLRPOPAIPIFSTVTGRLAEGDFNSYAMNRLOP  
 VSLAPVLTILAREDYTLFVLELBNMAVSLPVQOCFOAVKSAITLLSTLRGSEBHTL  
 RSLVGLFTIAGVTPAMNRLFPBGGSIVSLPSVPMOKERYASLSPISARLQILPAPGOV  
 LHGRVRSPLASLIPETDLSVKMPPLREHLIHLGALVYAGAAHVALLASALEIKGTA  
 PVLADIVFRPLVLEKEVENPVSVVSGEQGRSLRIQSPQSSPEVNTABEVL  
 PASSOVEGAQVALBELRARCPNGLEVSFVYKAMABEGIGLRFRQWEOALQOGGEA  
 LGRMAABASDQSPAPLHPGLIDSLFHLRAVPGIGRQDGVPIPIGIEGLCFIAP  
 RGVLANHVALEPESGDEGTLSADFLFESQKVAMIDGLLVKAPRAAPORHABD  
 LSEKLFQVWQSMPLAPAVBERGPMILVADRGAGEELARLKRGOVCHVLAAPESA  
 VTGELALQVBERALAEARGVYVLSLABEBSPOGSOALLEGSCAVOYLQALIK  
 LKRGTFPRMLVTRGQALLESBRPLAOLAPLMWGLAVRDEHPELRCTLLIDLPD  
 AADVGCLCFQELTTPREDRIAMRGARPAQRLSFLAVBERGTAPIISDGYVLLTGG  
 LEGICRCARMLVDAGARHLALIGRSGPEEARVSGALEAGGVTVRLOADVAARED  
 VKVLSALREVDMPPLRGVLAAGLDDALLVRODRRIQVLOPKVAGALNLTDEERG  
 DDLFEVLFESSIAAMGEGGQNYSAANAFILALAPAPAOGLPAHSLAGMWAVGM  
 VNAASRRAHLAALGIESIPDBGVAILLRLILRAASAHGITQIDMTTRQOQHAGHA  
 PLLAGLKKQQAAPSDOSSEEAADIOELIRAPBEEQKAEALFREVERIRIIDLPLQ  
 ADGRPLVLRFGFDSMLSLKLARLESTLGLSISAALKLSGLSLDKLVALLVIRIETLO  
 PTPVSTDDQMEBFRF"

gene  
 13537..17409  
 CDS  
 /gene="mtac"  
 13537..17409  
 /gene="mtac"  
 /note="peptide synthetase; peptide synthetase domains:  
 heterocyclisation, adenylation, PCP note: motif with weak  
 homology to putative NAD binding proteins from *Pyrococcus*  
 species located behind PCP; this motif may be involved in  
 thiazole formation from the proposed thiazoline  
 intermediate of the biosynthesis of myxothiazol"

/codon\_start=1  
 /transl\_table=11  
 /product="MtaC"

/protein\_id="AA019811.1"  
 /db\_xref="GI:6635396"  
 /translation="WIELLRPKAEVPPDAAROFEPPEPLTDIOETTFVGRGTFPGISGMS  
 CHLHEDLDRMDVARLSRAWQRLIEOHVMRLSVLPGEORVLESPPESIPLVDLR  
 QOALDLVEETKLAELRQOULSTSTSPGPPFTELATLIDBERTDIHPEDAITDADA  
 VLTLLLEEREETYLSDRPLKOVPSFRHVALEAARLTDSPROAEAWANRLNDMP  
 APALPVATEPRRLRPSFRRLSRLDAMARMLKERNANALITSGVCAFAEVLGA  
 MSEPRLTANTLRRRLRHPREVRLDGFPTVTLVRYANLDTGSPTAAVLRLOLAE  
 LDEHTRSGVWQRQAKMKGNAHPEVTSISLGHRSRPGGSGPFWLGEVYVAL  
 TQTPQMDLHVRREAGALYCHMOSPELPEGLLEGAPEAYITDLSLADNLAMLE  
 GTTVRLPEQAQARETFNATYASIPERLDGLFLAQOQORNAIIEPALMNSGEL  
 LGOAGSLAAHAEGARNEELTAVLEKMGQIVAVAIHLGASVLIPIDALPEER  
 RLILEGGVKTLLSHLEBSLSPASRYVAIADTAGAPRPPARASMAACTYIT  
 SCSGRPRGVKIEHRAANTLTDINERPGVAPRVASLSDGLSYVIDRGLSGLAG  
 AATVDPQEVDRPAHMLQVROQGVITWNSVPTLMENLVQVAREGLPEPTLRVLA  
 LSGWMIPTVLRDRRLRSRDRLVSLIGETASLWSIHLPIGRVDPARSIPYGRPML  
 NGRFVLEDEKLAPPDHVGGLYIGVGLAQGYRDEETTRAEFNNHRTGEIRYRG  
 DLGRFLPEGEIEFLGRKDPQVYKAGHRIEIGEIVAVLQRHPSIEAAVAAAGERSTR  
 LVAVLVPAGEGAPSEDEVLRSLFAALPQWVPAVPEVLEHLPLSSNGKVRKALPER  
 VLQSRASGLTDARGKTLASIMRLVSEPKRPSIEPPAGLLOLGATSVELIKLSML  
 EOEGRSPRTITELSLRNASEITAAVYERQPAEARKSDEAEVLPAGIRRDAGAVTA  
 QLEAPVEARQCYFKRVSHRTSAEVEVARVILRLSCSPALBESQLVQOSGAG  
 GYGVQVTLSTATGRVRLDGAAYVDEPLRHTLVQLGASALPALOSANAAVADA  
 AFSLYVGDREVIQRYGERMRDIALIEAGMAQLLETRAESDLGLCOLBELRFEGL  
 SKALRLBEHHVYLGVLGGGLAMEGAL"

gene  
 17406..27281  
 /gene="mtad"  
 17406..27281  
 /gene="mtad"  
 /note="PKS/NRPS; Mtad module 1: peptide synthetase  
 domains; heterocyclisation, adenylation, PCP note: motif  
 with weak homology to putative NAD binding proteins from  
*Pyrococcus* species located between adenylation core  
 regions A9 and A10; this motif may be involved in thiazole  
 formation from the proposed thiazole intermediate of the  
 biosynthesis of myxothiazol; similar to the motif found in  
 MtaC; Mtad module 2: polyketide synthase domains: KS, AT,  
 DH, KR, ACP"

/codon\_start=1  
 /transl\_table=11  
 /product="Mtad"  
 /protein\_id="AA019812.1"  
 /db\_xref="GI:6635397"  
 /translation="WSTSEFIVYUQOKIEIMBQDTRFPAPRQHLTPELTALDKD  
 KEQLEHRSRBSRANLOTVASPPDASRAPLDTLOINWYVWGROGAPAGVAAH  
 KYELLEHRSRBSRANLOTVASPPDASRAPLDTLOINWYVWGROGAPAGVAAH  
 SELEAPFTDLPRLSTLQVLPTRPFLFDRVAKLAGRVYHLHISPLTLADARFVO  
 PASQIEAHFEALRELSHQVLPTRPFLFDRVAKLAGRVYHLHISPLTLADARFVO  
 ILIEQASLYNPEAOVLPFRKTRDVCVLAAGRADPESRATONSILAWNRLELPA  
 APDLPLVNDTVKQCHQVFRKAGIDQSMRPAFNHAKAGITASMAKNAVGELTRA  
 YSRNRLTUNTLTFNRLPLIDVRYVGDPTFSGILLEDGKPSFARARLQGFQI  
 NDLEHSANSVKVLREARLRIGRWAGTAMPVPTSLISETGRSLSVSGRVKIVDVISQ  
 TPQVWLDQVLELDGGIYISDSDALFPAPALITAVPESYRLLHRLSSDASNDALS  
 RQPLPLPAOLARAEVNAATERVSOERIESLFLQAEOPGAVAAYAGQVLSYGLER  
 RSARLAPALAGVPPDQVAIVAEKGVQVVAALATIRAGAAVPLSPSPERLHA  
 LLEERQADVLTQSLLESLSLWMPGPRRLSYDRDEFLAPRGAALPVQGNMLAVIYT

Query Match 22.3% Score 944.8; DB 1; Length 42603;  
 Best Local Similarity 54.1%; Pred. No. 3.5e-133; Indels 120; Gaps 12;  
 Matches 2269; Conservative 0; Mismatches 1807;

QY	14	AGCTTCTGAACGAGCTCGAGCAACCGAGCTATCAAGCTCGCGCGCCGATGCGGAGCGCTCC	73
Db	17419	AGTTCAATGCTCTATCTTCACGAGGCAATGACTGATCGAAGGAGCAACGCTGC	17478
QY	74	AGATACAGGCCCCCAAGAACGCTTGAACCGGAACTCTGCTCGAATCTCCGAGCA	133
Db	17479	GATTCGGCGCGCCCTCGGCACTCACGCCGAGCTCGGACGCTGGAAGGCAAGA	17538
QY	134	AAAGACAGATCTTGACGATCTCCCTCA-----GAGACTCCCGGAGAAAT	178
Db	17539	AGGACAGCTCTTGAGAGATCTGCGTTCCGAGCGCGCTCCGCGGATCTTCAGACCGTGG	17598
QY	179	CCATGTCGCCGCGCCGAGCGGAGCGGACGCTCCGTTCTCTCTCAACAGATCCAAAGAT	238



Db 17599 CGAGCCCCCGGATGCGGCTGACCGCTTCCGCTTCCGCTCAACCAATCCGAATG 17658  
Qy 239 CCTATGCTGGGGCGGACAGGAGCGT---TAGGCTCCCAAGGGGATCCAGCGCTATC 295  
Db 17659 CCTATGCTGGGGAAGGAGAGGCTGCTTTCATGCGGGCGGCTCCAGCGCAAGTTACC 17718  
Qy 296 GCGAATACGACTGTACGATCTGACGTCGCGAGGCTGAGCGCGCTTTCGGAATGCG 355  
Db 17719 TCGAGCTGCGGTTCGACACGCTCGATCTGCGCGCTTTCGGAAGCATCTTTCAGCGGCTCA 17778  
Qy 356 TCGCGCGGACGACATGCTTGGGGCCACAGCTGCCCCGACATGATGCAAGTATCGAGC 415  
Db 17779 TCGAGTACACGACATGCTCCGCAATGTCGCTCCACGCGGACGAGGATGATGCTGG 17838  
Qy 416 CTAAGTCGACGCGGACATCGAGATC---ATCGATCTGGGGGCTTCGACGCGAGCACAC 472  
Db 17839 AGAAGTCCCGGCTTACCTTCACTCAGCGGTATGACTTGTGGGGCGCCCGCACTCAGA 17898  
Qy 473 GGGAGCGAGGCTCGTTCGTTTCGAGATGCGATGTCGACCGCATCTTATGACACCGAGC 532  
Db 17899 TCGAGGCGCATTTTCAGGCGCATTCGCGGGGAATCTCCACAGAGTGTCTCCAGCGAAC 17958  
Qy 533 GCCCTCGCTCTATCAGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 592  
Db 17959 GATTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18018  
Qy 593 TCAGATGATCTCTTAAAGTTCATAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652  
Db 18019 TCAGCTGATCTCTTCAAGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18078  
Qy 653 TCAGCTTTCAGAGAGATCCCGAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 712  
Db 18079 CTTCGCTTACCTGAATTCGGAAGGCAATGTCCTTTCAGGCGCATCTTCCGGAATT 18138  
Qy 713 A-----TGACTGCGCTGAGATTCGGAAGATCTGAGGCGCATCAAGATGATG 766  
Db 18139 ACTGCGCTGCGCGCCGCGAGCGCTGACCCGAGTTCCGCGGCTCAACAGAACTGCTTG 18198  
Qy 767 ATTAATGAGAGGCGGATGCGCGGCTCCACCTCCGCGGACGCTGCTGCTGCTGCTGCTGCTG 826  
Db 18199 CGTATGCGGAAACCGCTCTGAGAAATGTCCTCAACGCGCCCGATGCTGCTGCTGCTGCTG 18258  
Qy 827 ATCCATCTACCTTGAAGAGATCCGCTTCCGCAACAGAGATGCTGCTGCTGCTGCTGCTG 886  
Db 18259 AACAAGAGTGAAGGGGAGACACGATTCGTCGCGGAGAGGAGGATTCGATCAAGCT 18318  
Qy 887 CTTGGGTCGATGTAAGCGGCGCTGTCGAGGAGCGGCGCTGACCCGACGCGGCTGATCC 946  
Db 18319 CTTGGCGGCAATTCGCAATCATGCAAGGCGCGGGGATCACGGCTGCTGCTGCTGCTGCTG 18378  
Qy 947 TGGCTGCTTTTCCAGATGATCGGCGCTGAGAGGCGAGCCCGGCTTTCAGCTCAACA 1006  
Db 18379 GCGCGCTTATGAGGAAACGCTGCGGCTTCAAGCGGAGCAACGCGCTTCAAGCTCAACC 18438  
Qy 1007 TAAAGCTTTCACCGGCTCCCGCTCATCCGCGGCTGAAAGATTCACCGGAGACTTCA 1066  
Db 18439 TGAACCTTTCACCGGCTCCCGCTCATCCGCGGCTGAAAGATTCGCGGAGACTTCA 18498  
Qy 1067 CGTGAATGCTCTCTGAGATCGACACACTCGGAGCAAGCTTTCGAAACGCGGCTA 1126  
Db 18499 CTTGGGAAATCTGCTGAACTGATGCTGAAACCGGAGCTTTCGCGCGCGGCGGCG 18558  
Qy 1127 AGCGATTCAGAGAGCTGTCGGAAGAGATGATCATCTGCGAGCTAAGCGGTATCGAGG 1186  
Db 18559 GAGAGCTTCAGGAGAGTTCATCAATGATCTGAGACACTCGCTGTTTCGACGCGGAAGG 18618  
Qy 1187 TCACAGAGAGCGCGCGGCTCTGGGATTCACAGAGCGCATTTGTTCCCGGCTGCG 1246  
Db 18619 TGTCTCGGAGAGCATTCGCTTGGCGCTGGGATTCGCGGAGCGCATGCTTATGCT 18678  
Qy 1247 TCACAGAGCGCTTAAACAGCAAGTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1306

Db 18679 TCACAGAGCTTATCTCGAGAGACCGGCGCTCATTTGAGCGTCTCGAGGG---CGTGAAGA 18735  
Qy 1307 CGGTATCACACGAGCAGACCTCTCAGCTGCTGATGATCATGAGCTTACAGGACG 1366  
Db 18736 TGTTCGAGCTATCTTCCAGAGCGCCAGGTCGAGCTTCGACACAGGTTTTCAGAGCTG 18795  
Qy 1367 ATGGGAGCTTCGCTCGCGTGGAGATCTGACGAGAGTTCCTCCGCGCACTTCTG 1426  
Db 18796 ATGGCGGATCTATTAAGCTGGGATTCGCTGATTCGCTTTCGCGCGGCGCTGATCG 18855  
Qy 1427 ACAGATGCTGAAGCGTACGCTTTTTCGCGCGGCTTCACTGAGGAACA-----T 1480  
Db 18856 ACGCTGCTTCGAGAGCTGAGAGGCTCTTCAACCGCTTTCAGAGATGCTTCGCGGT 18915  
Qy 1481 GGGGTGAACAGGTGCTGCTTTCGCTTCGCTGCGCTGCGCACTGAGAGCGGCGAGCA 1540  
Db 18916 GGGATGCACTTCAGAGCAGCGCTTTCGCGCGGCTTTCGAGAGCGGCTTCGAGTCA 18975  
Qy 1541 ACGGACCAACGCGCTGTCGAGAGCATGCTGCAAGGCTTTCGCGCGGCTGCG 1600  
Db 18976 ACGGACCAACGCGCTGTCGAGAGCGAGTGAATGCTTCTTCTGTCAGGCGG 19035  
Qy 1601 AGCAGCTGCTTCATGACGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1660  
Db 19036 AGCTTCAACCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19095  
Qy 1661 TTTGCGCGCTTTCGCGGAGCTTTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1720  
Db 19096 TCGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19155  
Qy 1721 CATTTGCTGCGGCTGATGAGAGAGCTGCGAGCAAGTTCGCGGCTTTCGCGGCTG 1780  
Db 19156 AGCTGCTGCGGATGCTTTCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 19215  
Qy 1781 TCGAGTCAAGCGCGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1840  
Db 19216 TTTGCGCTGAGAGCGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19275  
Qy 1841 ACCTCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1900  
Db 19276 CTTGCTGAGAGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 19335  
Qy 1901 TGTTCATGCGCGCGGAGATTCAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1960  
Db 19336 TGGATGCGCGGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19389  
Qy 1961 GCGACACGCTTCGATGATGCTTCAAGACACTTGGAGTCTGCGGATTCATCTTCA 2020  
Db 19390 AAGCCCGAGAGGCGGCTGCTGCGGCTTCAAGGCAACATCTGCTTCACTTCACTTCA 19449  
Qy 2021 CTTGCGGATTCAGAGGCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2080  
Db 19450 CTTGCGGCTGAGAGGCGGCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 19509  
Qy 2081 CCATCTGAGCATCAAGAGCGCTTTCGAAATAGGCGCGGAGAGAGAGGAGGAGGAGGAGGAGG 2140  
Db 19510 CCGCTCTGAGCATTAAGAGGCTTTCATATGAGGCGCGGAGAGGAGGAGGAGGAGGAGGAGG 19569  
Qy 2141 CTTGCTGAGCTTTCATCTCTGCTTATGATGCTTTCGAGGATCTTTCGCGGCGGAGGAGG 2200  
Db 19570 CGTCACTGAGCTTTCATCTCTGCTTTCATGAGCCTTTCGAGGATCTTTCGCGGCGGAGG 19629  
Qy 2201 CGATGCTGAGGCGGAGGCGGCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2260  
Db 19630 CGATGCTGATTCGAGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 19689  
Qy 2261 AACGAGAGAGGATGAGGCTGCTGAGTTCGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGG 2320  
Db 19690 AGAAGAGGAGGATGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 19749  
Qy 2321 ATTTGAGGCTGCGCGGAGTTCGCTGCTGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2380  
Db 19750 TCTCGAGGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19809



QY 2381 ACTGATCCCGGTGGGCTCTGCTGGAGACTCCAGGCAATCAAGCCCGGCTGTCTGTGA 2440  
 Db 19810 ACTGATTCCTGGCTCAACGCTGCGCCGGGATCCGTCCTGTCCAAAGAACGCCGAGCTGG 19869  
 QY 2441 TCAGCTGGGCGGGGACCCGAAGCGTCGATCTGTGTCATCGGATACCCCGTGAAGAAC 2500  
 Db 19870 TCAGCTGGGTCGGGACCGGAGGCTTCATCTGTGTCATCTCTACCGGATTTGGGAGTG 19929  
 QY 2501 TCAGCTTATGTTGGGAGAGCATCCCTACGCGGCTGTGCGGCAACGAGCGTTTCACG 2560  
 Db 19930 TGGGGGCGGCTGGCGGCAACATTCGTCGTCGCGCCATGTGATCAAGCCCTTCTATG 19989  
 QY 2561 TCGTGAATAGGCGCTCTGAACCGCGCCGCTGTGGTTCCGGGCAACTCTACATTTGGG 2620  
 Db 19990 TCGTGAATAGGCGCTCTGAACCGCTGTGCGGATCGGCGCCAGATGTCATTTGGG 20049  
 QY 2621 GGGTCGGGCTGGCACTGGGCTACTGGCGGATGAGAGAACGCGCAAGAGCTTCTCG 2680  
 Db 20050 GCATTGGCTTCTGCTGCTGCTATTCACGGGATCCGCTGTGTCACGCGGAGCGCTTCATG 20109  
 QY 2681 TGCACCCCGAGACCGGGAGCGCTTACAGAACCGGCGATCTGGGCGCTACCTGCCG 2740  
 Db 20110 TCCATCCCAAGACCGGGAGCGCTGTACGCAACGGGTGACCTGGGCGGTTCTATGCCG 20169  
 QY 2741 ATGGAACATCGAGTTATGCGGCGCTGAGAGCAACCAATCAAGCTTCCGGGATACCGG 2800  
 Db 20170 AGGGGAACATCGAGTTCTCTGAGACGGGAGACTTCCAGGTCAAGATCCAGGCTATCCGA 20229  
 QY 2801 TTGAGCTGGGGAATCGAGAAACGCTCAAGTCGATCCGAACGTCAGCGGCTGA 2860  
 Db 20230 TCGAGCTGGGGAATCGAGCGCGGCTGATCTCGATCAGCGGCTCTCTCGGTTGG 20289  
 QY 2861 TTGTGCGCTGGGAAAGACGCGGCAACAGCTCTTCTAGCTTATGTGTCCGGAG 2920  
 Db 20290 TCAATGCCCTTGGAAAGCCCGGGGAGACTCGGCGCTCGTGTGCTGTCTCCCGAGG 20349  
 QY 2921 GCAACGAGAGACCGCTGCGGAGCAAGAGCGGAGCTCAAGACGAGC---GAGTCGAG 2977  
 Db 20350 AATTGACGTGAATGCTCCCGGCAAGCGCGCTCTCGGCGAGCAAGCTCTCTTGGAGT 20409  
 QY 2978 CGAGAGCAACACCGCGGAGCGAGCGCTTGAAGCGAGCGAGAGG----- 3025  
 Db 20410 CGGCTCCGAAGCTTGCCTGAGAGTACAGCGCTCTGTGAGAGCCAGGGGCTGGGCTGATCA 20469  
 QY 3026 -----TCAATTCAAGCTGCTGACACGAGCTCTCGGAGGAGCTGG 3067  
 Db 20470 CCGATACGCTGGAACGCTGAGATTCAAGCTGAAGCAACAACTGCGCGGATGACG 20529  
 QY 3068 AGGGAAGCGCGCTGCTGATCTGACCGGAGAGATCCGGGAGGCGGGGCTGAGCTCT 3127  
 Db 20530 GAGCGCGGCTCAAGTTGCTGCTCAAGAACCGGAGTTGATGATGAGGCTTCCGCAAGAGT 20589  
 QY 3128 AGCGCGTGGCGGTCGCTGAGCAACCTTCTTGAAGGCGCGGATTCGTTTGTGAGTTTG 3187  
 Db 20590 ACATGAGGCGGAGAGCGGCGGAGCTTCTCTCCGAGCGGCTCTCTGAGAGCGGCTCA 20649  
 QY 3188 GTGATTTCTGAGCTGCTTGAAGAGCGTGAAGCGGCGGAGCCCTTCCCAATTCC 3247  
 Db 20650 GCGAGTTGCTGAGCTGTGATGACAGATCGCTCTCGAAGATCTCAATGATGCCCAAGTAC 20709  
 QY 3248 GTTATTCATCGGCGGAGCAAGTACCGGTCGAAACCTACGCTATGTCGAATCCGGGCG 3307  
 Db 20710 GGTATGATCGGCGGTCGCTTATCCGCTGAGGTCATCACTGACGTCGAAGCCAGGGCG 20769  
 QY 3308 GCATCGAGG---CGTGAAGAGGCTTCTATTTATACACCGCTTGAAGACCGTTTTCG 3364  
 Db 20770 GTGTGGGGGGGCTCGGGGCGGAGCTACTATTTATACCGGAGCTCAAGAGCTGTGCG 20829  
 QY 3365 TGAAGCTTCCGATCAAGGATCGAGCGGAGCGGAGCTTGGCAAACTTGAAGCTGT 3424  
 Db 20830 TCGTACCGCGGATGCGGCAATGAGACGAGCCAAACGCGTCCGGAACCGGCGGCTCT 20889

QY 3425 TCGATGAAGCGGCGTTCAACCTCCCTGTTGTTGGGAGGATCGAGCCGATCGAGTCTGT 3484  
 Db 20890 TCGATGCCGCGCGCTTCTCCGTTTCTCTGTTTGAAGACTAGTGCATTCGCGCGCTGT 20949  
 QY 3485 ATGATTCGTCGCGAGAAATTTTTCCTGTGAGAGCGGATATATGCGGAGCTCTGA 3544  
 Db 20950 ACGGAGATAGGCGGGGATTTTCTGATCTGAGAGCGGGGTCATATGCTCAGCTCTCA 21009  
 QY 3545 TGAAGCAGCGCTTCTCTGCAACATCGGCTGTCTCGGTGGGCAATTCATTTTGAAC 3604  
 Db 21010 TGAAGCTTCCCGCCCAAGAGGTTTGTGCTCCATCGGCTGATGAGACTTCGAGC 21069  
 QY 3605 AGTTGCGCGGTTTTCGACCGCGAGCATTCGAGCTTACGTGACGAGCATGTGGGCG 3664  
 Db 21070 CCTGCGCTCGAGCTCGCGTTGAGAGCAACACAGTCTTCTGCAAGCTTCTGTGGGG 21129  
 QY 3665 GCGGCTGAGACCGCGGAGTTCCAGGTCTGATGACCTGTGAGATTCCTACCGAGG 3724  
 Db 21130 GCGGCGAG-----TGACGGCGG 21148  
 QY 3725 GCGGCAAGACGCGGCGGCTCTCCGCGCGGAGCACTTCGCGATATCTTGGCG 3784  
 Db 21149 GCGTCGCGCAAGCCAGCAAGCCAGGCGGAGTGTCTTCTGAGAGG 21207  
 QY 3785 ACTTCTGAGGACCAACCTACCGGATCATGAGTCTTACAGTCTTGTGAGCTGATG 3844  
 Db 21208 GTTACTGAGCTGCAAGCTGCGGAGTACATGTTCTGCTCTGTTCTGTTGATTT 21267  
 QY 3845 CTTTCGCTGACGTCACAGCGGAGGTCGATGCTGAGGCTTGGCGGAGGAGATA 3904  
 Db 21268 CCTTCGCTGACGTCACAGCGGAGGTCGATGCTGAGGCTTGTGCTCGGCGGAG 21327  
 QY 3905 CCGTGTCCCGCGGATTCGGGCGACAGCGGCGGAGGCTTGGAGAGATCTTCG 3964  
 Db 21328 TGTGTAAGTCAAGCGCGGCGCTGCGCGGAGGCTTGGAGAGATCTTCG 21387  
 QY 3965 TCGCGGCTGTAAGGAGTGTGCGGCTGAGAGTGTGCGGCTCCAGAGAGCTTCTG 4024  
 Db 21388 CCGCATCTTCAAGAGTGTCTCAACTGCAAGAGTGTGCAACCGCACTTCTTCG 21447  
 QY 4025 ATCTTGTGCAATCATCATGATGTCATGATGAGAGGCTTGTGAGAGAGCTTG 4084  
 Db 21448 ATCTGGGGGAGACGTCGCTTCAATGTCAGATGATGATGAGAGAGGCGGCTCA 21507  
 QY 4085 ATAGGAGATGCCATCAACCGAGTGTTCAGTACCGCAACTCGGCTCTGCTGGG 4140  
 Db 21508 AGGCGGATATCCGATTCGCGAGATGTTCCGATACCAACCGTCAAGTGCCTGGG 21563

RESULT 11  
 AF210249 77457 bp DNA linear BCT 29-AUG-2000  
 LOCUS AF210249 Streptomyces verticillius bleomycin biosynthetic gene cluster.  
 DEFINITION complete cds.  
 ACCESSION AF210249 AF149091  
 VERSION AF210249.1 GI:9937210  
 KEYWORDS Streptomyces verticillius  
 SOURCE Streptomyces verticillius  
 ORGANISM Streptomyces verticillius  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycinae; Streptomycetaceae; Streptomyces.  
 REFERENCE 1 (bases 74421 to 74693)  
 Du, L. and Shen, B.  
 Identification and characterization of a type II peptidyl carrier  
 protein from the bleomycin producer Streptomyces verticillius ATCC  
 15003  
 Chem. Biol. 6 (8), 507-517 (1999)  
 JOURNAL MEDLINE 99352421  
 PUBMED 10421758  
 REFERENCE 2 (bases 1 to 77457)  
 Du, L., Chen, M., Sanchez, C. and Shen, B.  
 An oxidation domain in the BlnII non-ribosomal peptide synthetase  
 probably catalyzing thiazole formation in the biosynthesis of the

anti-tumor drug bleomycin in *Streptomyces verticillius* ATCC15003  
JOURNAL  
FEMS Microbiol. Lett. 189 (2), 171-175 (2000)  
MEDLINE  
20389599  
REFERENCE  
PUBMED  
10930733  
AUTHORS  
TITLE  
3 (bases 1 to 77457)  
Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.  
The bioynthetic gene cluster for the antitumor drug bleomycin from  
*Streptomyces verticillius* ATCC15003 supporting functional  
interactions between nonribosomal peptide synthetases and a  
polyketide synthase  
JOURNAL  
Chem. Biol. 7 (8), 623-642 (2000)  
MEDLINE  
20500448  
PUBMED  
11048953  
REFERENCE  
Du, L., Sanchez, C., Chen, M., Edwards, D.J. and Shen, B.  
AUTHORS  
TITLE  
Direct Submission  
JOURNAL  
Submitted (01-DEC-1999) Chemistry Department, University of  
California at Davis, One Shields Avenue, Davis, CA 95616, USA  
On Aug 29, 2000 this sequence version replaced gi:5326869.  
COMMENT  
FEATURES  
source  
Location/Qualifiers  
1..77457  
/organism="Streptomyces verticillius"  
/mol\_type="genomic DNA"  
/strain="ATCC15003"  
/db\_xref="ATCC:15003"  
/db\_xref="taxon:29309"  
complement (1199..2185)  
/codon\_start=1  
/transl\_table=1  
/product="syrp-like protein"  
/protein\_id="AAG02342.1"  
/db\_xref="GI:9937211"  
/translation="MTGSGVTLPLGGIIPRPRGEGLTGAEVLDGLDAGDPWRAH  
GPRLEBRLATGTLTGLHGLPRDGGVGFHVDVSGDPLPYTERSPRSPVGNVY  
TSTERYADQPIPMHNENSYAAHWSTLTFCHTAPDTGATPIADGRAVLDLIRAEYR  
RTSGGVVTRTFRADWGLSWQEARQTEDEKRDVERHCAHQEFSWDQDVRTNHRP  
ATAVDPGAVWVFNQALFHPSSLDPLRQVLETYGENGLPRDALFADTPIPDAD  
LATRAATRAALALPMREGDIMLVNIRMAHGREPFGERRVLVAMTSADS"  
complement (2182..3918)  
/codon\_start=1  
/transl\_table=1  
/product="peptide synthetase NRPS12"  
/protein\_id="AAG02343.1"  
/db\_xref="GI:9937212"  
/translation="MTLTLRDAFLDQAARTPDAAVYHGDVTWYRLELRAGMART  
LAARGGCTLVARLPGRPPEVAALAVLTGAGVYPLADDDPDRRHHLLDCAAA  
LLAHPRDRGRTLTPDBALAPAPFPDAPVRAQDPAYVITSSSGSKGLVBOGA  
LGAYLAQARARYDGLSGRTVLHSLSFMAVTSIMGPLVSGAHLVLDLKAISGTOP  
PPAASARSPFLKVTPLSLPLGLPDSCLPTGQVITGGEALTGSLGPMRAHDPVY  
VNEYGPTGATVGCAYTVRPGDAVDPGAPIGRPAPGATRLVLDADGEPVAVGVEL  
HLAGDOLARGVLRPLTEERFVDPDPFAADGSRMVRGDLVREERPDGLLEYLGRADGO  
VKVSGYRTERGEIENVLGHAGVRCANVANGEDARLVAVVYVDPDPSPGTAAAPR  
HAERALPYMPATFVTVELPLTPNGLDLDDALGPPAGAGPDRTPAETLLCELL  
ARALGIPEDIDADADFLTSGSITALKLVAGARRVRLLETLVLRERTVRIIAAOP  
DAASPLAEVPE"  
complement (3915..4733)  
/codon\_start=1  
/evidence-not experimental  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAG02344.1"  
/db\_xref="GI:9937213"  
/translation="MSKVAHMSRI RTGDAPGVLANVDFYGTGROEATFRLCDLTD  
PVEYHAYPAPADGDMSTATGAGLRMTBGLDVTYLAQRPRALVGYAGGVPSALAA  
DALYREGRHPRVVLFPNSAPGVALITRDFGLTAGMDLTDGEAALATATYIRRA  
HAPALVVARAYALYREGCDLCEKRGVDSFGAEALAAVLSYALVLTALVLPPT  
PLMRGASVLTSGTDFDVEHGFVARELLSSPQVVALTALREHBSR"  
complement (4741..6024)  
/codon\_start=1  
/transl\_table=1  
/product="putative transporter"  
/protein\_id="AAG02345.1"  
/db\_xref="GI:9937214"

/translation="MTARVADATPTYLAVLAAREAPRLGSCLRKSPAVLPLALLS  
VRDATGSPFAVAGSGALSAATLTFAPARALIDRRSSGSLVLTPTVYLGLVLT  
LAEAPETPALVLAAGVAFAPLPGCTMRVLMARILHGRPLHVTVAADSVLEVY  
FTVGBLLAGLIVAAFLASMITWTVLIAGTACFVLSAATAAPAGDEDPHGR  
PMALPKRRTVLSFGVGLVGVVQVLPFIADNAGSPGAGITLSLHSGASANGCLA  
YGRITAMSTVRRFVVLVGTFTLAVLCLCTPASFVPAGARLVLGLCLAPLFTYALV  
VNDVLTSGAPLEANTMVSSTANNAGSAAGVLLDSKGPVTVTVAFAVAATA  
VMVLTARTTLLGHPPEAAATFADRTAAEAEE"  
complement (6021..6773)  
/codon\_start=1  
/transl\_table=1  
/product="thioesterase"  
/protein\_id="AAG02346.1"  
/db\_xref="GI:9937215"  
/translation="MDPLTRVNMPSFGCDGRRPRVLCALPYAGTAAPKDPAPAL  
PGVEILTALHPRGDRFTEPRPAETLETERACEALPQGLPTVYVGHSGALLGYE  
VVARLARGAENLILAAKCRPPHVPDPDAPVTAEALATLRERPMPTDTRBEEIM  
EAVPLAVLADITAGDRYHRPRPLDPLKVIYIGADDDGTWRTTLGACTARDCV  
VVLPGHYFLETDRAAVLTVAATDLAEAEVGA"  
complement (6784..7797)  
/codon\_start=1  
/transl\_table=1  
/product="clavaminate acid synthase-like protein"  
/protein\_id="AAG02347.1"  
/db\_xref="GI:9937216"  
/translation="MTDLPLRTVVALTGEESAENVDDLRTTLADVPEVSTVGLHRTTLA  
AGELPLIRIAELTCMRLYDSPRALVVTGPGVDERIGTPPARAPAPBERTRDEL  
LHAAALGGEAPGATQONGRLVHDLVVPBEETQSSSETELLMTEDEPHLRCD  
YVGLCLRNHQRATTVGMFDLSRLTTEDDEPVLVWSPYLRPTSHPTQNAQTRSA  
ERFPAIAEMDPAERVAVLFGDEPDPLRLDPAVAPSPAGPAARARVDTYALIDE  
LRHVLDAGSLLVADVNYQAVHGRKPFPAAYVADGRDRLKRVNITDRLRRSRRSATS  
LV"  
complement (7859..9214)  
/codon\_start=1  
/transl\_table=1  
/product="putative transporter"  
/protein\_id="AAG02348.1"  
/db\_xref="GI:9937217"  
/translation="WMKSRLRDLGSEDPVVAQESPODAGPTPCGDDGANYFAAL  
AALLVEVPVRLPHNAGLGRVHSHSOAAVAGPNOVASATGVGADGGHAGRGO  
VTVGVGREVAGGDDDELGAEDRPHARLRDLGLMAERLADLLVNPLOTVVGODL  
RGQVSHDGLGDLVPGQGLSLGSLQRRRDGVGTALASVGPQRQGPAAASCRG  
LVARQODQRALRAVVEGPPQDQACQARVAETDHPDVPVGHETVGGQOGEVGL  
GHVDRREVSVAAGFDDVAGVGRALPVSAGHVDRAAGVDGLTAGRREGQO  
ORGGRAQDVHCPVDLVGVGVEDLTGCGDRLIVADLRPHQRTLVHRRVMPMLARY  
DPGDDPPSLAHSLTLPSNIPSTRGTPRCHLRKCRKRTKSQSAARAPRRGRLLVS  
H"  
complement (9211..17736)  
/codon\_start=1  
/transl\_table=1  
/product="peptide synthetase NRPS11-10"  
/protein\_id="AAG02349.1"  
/db\_xref="GI:9937218"  
/translation="MASDALSAHLEVSDCRQGGITVTNDGSELGAGNVAAVREERY  
AIAPERTAIIHKGATGYDELRRAELTARLADAGAGPTLVVVALPRDVLATLTC  
ALKEGAACLPDGLPAGRLREIMADASPDVLTPTAVVAPFGDPLADAPPT  
CSAVLPGRSAGTASEIAYVLYPTTPDKSSNSVSYDMARVLDPPPARPARBETIR  
LVAPLISGRLVLDADETPRPRPRPRMVEVVAQWCAVYGVRCGRFPLD  
GKSLAAVQVVARLRKLVGLVPLRALFDLPALVYBELARARABAGGQVREBALEP  
VGRSEPLPLSFAQORLWFLDRLMEDRAFYMCDAFRVGGIDIGLARALRMVLGRHE  
TLRTAFPERDGVQPOLVGPADGPARVAAPTRVDSLLEBAEEREAVRNIVAAEART  
PFRPDAGALLRVVVALDDDDVLTSTHIIIVSAMVGVLDLGLRYRECVGDP  
ALPPADQVADPAVWQAMAGPVQSEHLLVWKRALDGAQSVLRLPMDPRPVAOSER  
GETVGRPLPDLVAALEKLRBQATYPTTLTGAFQVILARHAGQEDIVGVVPAAGRT  
RTEPELVGFVNTVLPKATCAPGLSFRDLDDVREALAGAFHODLPEFVLEALAP  
ERIDGNHPLQVVFQLLGATPAARDLGTEVERAPVEAVSQFSLDILKADDSYR  
GLIAYCPDLDRRMEVLYGVYTLTGAADADPRPGELPLSGARLRLLDGFGK  
DAAYAGSGVPERFAEVARTAPARATCATTLTPLELNDRYERLRAQLILGAGTSE  
TPVAVVLRPSTDSVALLAVMRAGGVVPLDPMPARATYIILDDTAAVYTRDPLA  
LPGRLHDPRRPADGIVPARIDPDQAAVITYTSSGTGAPKQVYVGHRSNHTSL  
QATPLGDPYLAGADGVPGDQAKRITLLAPFTFDASMEQSNMLAGHELTVEDER  
RDPBALVRPVERHRIDIVITTSQLELHSHGLDGMASPMVWVGEAAVSPSLMRTL  
RDGRRTRCFNLVYGTETATVDATCHDLSDDPADVPVITGTLPHTHVRVLDRRLRPVGV

CDS  
/codon\_start=1  
complement(18307..18660)

Query Match 17.2%; Score 728; DB 1; Length 77457;  
Best Local Similarity 55.9%; Pred. No. 3.8e-105;  
Matches 1545; Conservative 0; Mismatches 1185; Indels 36; Gaps 7;

182 TCGTCCCGCCCGCCGAGCGGCGACGCTCCGTTCTCTCAACAGATCCAAAGATCTT 241  
Db TCGTCCCGCCCGCCGAGCGGCGACGCTCCGTTCTCTCAACAGATCCAAAGATCTT 241  
61043 TCGTCCCGCCCGCCGAGCGGCGACGCTCCGTTCTCTCAACAGATCCAAAGATCTT 241  
242 ACTGCTGGGCGGAGAGAGCGTTTACGATCCCGCCGAGG---ATCCAGCGCTATCCG 298  
Db ACTGCTGGGCGGAGAGAGCGTTTACGATCCCGCCGAGG---ATCCAGCGCTATCCG 298  
61103 ACTGCTGGGCGGAGAGAGCGTTTACGATCCCGCCGAGG---ATCCAGCGCTATCCG 298  
299 AATGAGCTGTACGATCTGACGTCGAGCGGCTGAGCGCGCTTTGCGAAAGTCGTCG 358  
Db AATGAGCTGTACGATCTGACGTCGAGCGGCTGAGCGCGCTTTGCGAAAGTCGTCG 358  
61163 AATGAGCTGTACGATCTGACGTCGAGCGGCTGAGCGCGCTTTGCGAAAGTCGTCG 358  
359 CGCGGCGACGACATGCTGGGCGGCGACGCTCCCGACATGATGAGAGGTCGATCCGTA 418  
Db CGCGGCGACGACATGCTGGGCGGCGACGCTCCCGACATGATGAGAGGTCGATCCGTA 418  
61223 CGCGGCGACGACATGCTGGGCGGCGACGCTCCCGACATGATGAGAGGTCGATCCGTA 418  
419 AAGTCGAGCG---CGACATCGAGATCATGCTGCGCGGAGCTCGACCGGAGACAGCGG 475  
Db AAGTCGAGCG---CGACATCGAGATCATGCTGCGCGGAGCTCGACCGGAGACAGCGG 475  
61283 AAGTCGAGCG---CGACATCGAGATCATGCTGCGCGGAGCTCGACCGGAGACAGCGG 475  
476 AAGCGAGGCTGCTGCTGCTGCGAGATGCGATGTCGACCGCATCTATGACACCGAGCGCC 535  
Db AAGCGAGGCTGCTGCTGCTGCGAGATGCGATGTCGACCGCATCTATGACACCGAGCGCC 535  
61343 AAGCGAGGCTGCTGCTGCTGCGAGATGCGATGTCGACCGCATCTATGACACCGAGCGCC 535  
536 CTCGCGCTCATACGCTGCTGCTGCTGCGAGATGCGATGTCGACCGCATCTATGACACCG 595  
Db CTCGCGCTCATACGCTGCTGCTGCTGCGAGATGCGATGTCGACCGCATCTATGACACCG 595  
61403 GGCCTCTGTT---CGACGCTGCTGCTGCTGCTGCGAGATGTCGACCGCATCTATGACACCG 61459  
596 GTATCGATCTCATTAAGCTTACCTAGAGAGCGCTGTCATCTTCAAGAGACTGAGCTCA 655  
Db GTATCGATCTCATTAAGCTTACCTAGAGAGCGCTGTCATCTTCAAGAGACTGAGCTCA 655  
61460 GGTGAGAGCGGCTGATCTGCGACGCGCATGCTGAGCTTCTGCTGCGCGGAGCTGAGCT 61519  
656 GCTTCTACGAGATCCGAGAGCTCTCTCTGCTGCTGAGAGCTTCTGATCCGCGATTATG 715  
Db GCTTCTACGAGATCCGAGAGCTCTCTCTGCTGCTGAGAGCTTCTGATCCGCGATTATG 715  
61520 CCGCTTACGAGATCCGAGAGCTCTCTCTGCTGCTGAGAGCTTCTGATCCGCGATTATG 715

716 TACTCGCGCTGAGAGCTGCGAAGAGTCTGAGCGGCTCAACGATGATGATCTGAG 775  
Db TACTCGCGCTGAGAGCTGCGAAGAGTCTGAGCGGCTCAACGATGATGATCTGAG 775  
61580 TCTCTCATGAGAGCGCTCCGCGGAGATCCCGGATGACCGGCGGAGCGGATCTGAG 835  
776 AGCGCGAGATGCGCGAGCTCCGAGCTCCGCGGAGCTTCCGATGAGAGCGGATCTGAG 835  
Db AGCGCGAGATGCGCGAGCTCCGAGCTCCGCGGAGCTTCCGATGAGAGCGGATCTGAG 835  
61640 GCGAGAGCTGCGAGCTGCGCGGAGCTCCGCGGAGCTCCGCGGAGCTCCGCGGAGCT 895  
836 CCGTGAAGAGATGCGCTCCGCGAGACGAGAGCAATGCTGCTGCGAGCTTCCGAGCT 895  
Db CCGTGAAGAGATGCGCTCCGCGAGACGAGAGCAATGCTGCTGCGAGCTTCCGAGCT 895  
61700 CCGTGAAGAGATGCGCTCCGCGAGACGAGAGCAATGCTGCTGCGAGCTTCCGAGCT 955  
896 GATTGAAGAGATGCGCTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955  
Db GATTGAAGAGATGCGCTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955  
61760 GCGTGAAGAGATGCGCTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955  
956 TTTCCGAGATGATGCGGAGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1015  
Db TTTCCGAGATGATGCGGAGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1015  
61820 TCGCGAGATGATGAG 1075  
1016 TCAACCGCTCCGCTCCATCCGCGCTGAGAGAGATGATGAGAGAGAGAGAGAGAGAG 1075  
Db TCAACCGCTCCGCTCCATCCGCGCTGAGAGAGATGATGAGAGAGAGAGAGAGAGAG 1075  
61880 TCGACCGCGCGCTCCGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1135  
1076 TCGTCTGAGATGAG 1135  
Db TCGTCTGAGATGAG 1135  
61940 GCTGCTGAGAGATGAG 1187  
1136 AAGAGAGAGATGAG 1187  
Db AAGAGAGAGATGAG 1187  
62000 AGCGCGCTGAG 1240  
1188 -----CGAG 1240  
62060 AAGCGCGCTGAG 1241  
1241 TCGTCTGAGATGAG 1300  
Db TCGTCTGAGATGAG 1300  
62120 ACTGCTGAG 1360  
1301 GAACTCCGAGATGAG 1360  
Db GAACTCCGAGATGAG 1360  
62180 GAGAGCGCTGAG 1420  
1361 AGCAG 1420  
Db AGCAG 1420  
62240 AAG 1474  
1421 TCTGAG 1474  
Db TCTGAG 1474  
62300 CCGTGAAG 1534  
1475 AAG 1534  
Db AAG 1534  
62360 AAG 1594  
1535 GCGAG 1594  
Db GCGAG 1594  
62420 GCGAG 1654  
1595 GCGAG 1654  
Db GCGAG 1654  
62480 GCGAG 1659  
1655 AAG 1659  
Db AAG 1659  
62540 GCGAG 1714  
1715 GCGAG 1714  
Db GCGAG 1714  
62600 CCGAG 1774  
1775 CCGAG 1834

```

Db      62660 GCGTCTCGAGATCGGGGGCGGCGTACCTGCCCCGAGACCCGAACTGCCGCCGACAGCGGC 62719
Oy      1835 TCCACTACCTCTCGATCATGTGATGAGTAAGCTCTGCTGATCCGACGATCGATGATG 1894
Db      62720 TGTTCACCTCTGTAACGGGCGCCCGAAGCGCCCTCTCTCTTCAACGAAAGCGCCCTGTGCG 62779
Oy      1895 GCAACTGTCAATGAGCCCGCGGGGATCCAGCGGCTGCTGATGAGGAGCGCGCGCTCGAAG 1954
Db      62780 ACAAGCTGCGGCTCCCGCTGGCGTCAAC--CGTCTCGCGGTGAGACGACCGCGGCC 62836
Oy      1955 GCGAGCGGACAGCAGCTTCCGATGATGCCATTACAGACCTTCCGATTCGCGATGTCA 2014
Db      62837 TCGACGCGGAGCGGCGCGCGCTGCAAGCTGCAAGAACCTCAACGACCTGCGCTGACACA 62896
Oy      2015 TCTACACTCGGGATCCACAGAGGTTGGCCCAAGGGGGGTATGATGATGATGAGGGGTGCGG 2074
Db      62897 TCTTCACTCGGGCTCCACCGGCAACCAAGGGGCTGATGATGATGATGATGATGATGATGATG 62956
Oy      2075 TCAACACCATCTCTGACATCAACGAGCGCTTCAAAATAGGCGCGGAGACAGGGTGTGCG 2134
Db      62957 CCAACACCTCTGGAATGGGTCAACCGCGCTTCCGACCGGCGCCGCGGACCGCGGTCTCG 63016
Oy      2135 CGTCTCTCTGCTGAGCTTGCATCTCTGCTGATGATGATGATGATGATGATGATGATGATG 2194
Db      63017 CCGTCTCTCTCCCGAGCTTCACTCGCGCTGCTGATGATGATGATGATGATGATGATGATG 63076
Oy      2195 GCGGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2254
Db      63077 GCGGCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63136
Oy      2255 TGATCGAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2314
Db      63137 TGATCGGCGGAGAGGGGTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATG 63196
Oy      2315 TCGAGCATTTGAGGGTGTGCGCCGATGCTGCTGATGATGATGATGATGATGATGATGATG 2374
Db      63197 CCGAGTACGCGGAGCGCTCCCGCCCGAGCGCGCTGCGGACCTGCGGCGGTGCTCTCA 63256
Oy      2375 GCGGCGACTGATCCCGGTGGGCGCTGCTGCGGATGATGATGATGATGATGATGATGATGATG 2431
Db      63257 GCGGCGACTGATCCCGGTGGGCGCTGCTGCGGATGATGATGATGATGATGATGATGATGATG 63316
Oy      2432 TGTGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2491
Db      63317 CCACCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63376
Oy      2492 TGAGGAACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2551
Db      63377 TCGGGAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63436
Oy      2552 CGTTCACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2611
Db      63437 GGTGAGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63496
Oy      2612 ACATTTGCGGGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2671
Db      63497 ACATTTGCGGGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63556
Oy      2672 GCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2731
Db      63557 GCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63616
Oy      2732 ACCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2791
Db      63617 ACCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63676
Oy      2792 GATACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2851
Db      63677 GATTTCCGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63736
Oy      2852 ACGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2911

```

```

Db      63737 CCGGCGGCTGATCGCACCGGTGACCCGCGGGCGACCGCGCTGCTGCGCTGCGG 63796
Oy      2912 TCCCGG 2917
Db      63797 TACCGG 63802

RESULT 12
AL646066/c
LOCUS
DEFINITION
Ralstonia solanacearum GM11000 chromosome, complete sequence;
segment 10/19.
ACCESSION
AL646066 AL646052
VERSION
AL646066.1 GI:17428714
KEYWORDS
SOURCE
ORGANISM
Ralstonia solanacearum
Ralstonia solanacearum
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE
AUTHORS
1
Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
Arlat M., Billault A., Brotier P., Camus J.C., Catalioto L.,
Chandler M., Choisme N., Claudel-Renard C., Cunha S., Demange N.,
Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,
Sguier P., Thebaud P., Whalen M., Mincker P., Levy M.,
Weissenbach J. and Boucher C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Nature 415 (6871), 497-502 (2002)
TITLE
Ralstonia solanacearum
JOURNAL
MEDLINE
21681879
PUBMED
11823852
REFERENCE
2 (bases 1 to 189050)
Boucher C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
UMGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequece.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
1..189050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GM11000"
/db_xref="taxon:305"
170..3325
/gene="RSC1698"
/note="synonym: RS02885"
170...3325
/gene="RSC1698"
/function="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE BACTERIOPHAGE PROTEIN"
/protein_id="CAD15400.1"
/db_xref="GI:17428715"
/db_xref="SPTREMBL:O8XXR0"
/translation="MKNTIGGGKGGSSSTTEAPDSLSHYARVLDVSEGEIA
GLVNGLOSITLSTPLANADGTLNFQNVADVYPGSDODGATGFPBSEVSESNVGEV
LATMPVRAIANTOLSAVRVOLSVPLSRADTSSGNGINGRYEVAIDLAVDGGALDOV

```

1ATAFDGKTSKVTBTRHELPAKTKGWTIRVRITPANSGTIDAVTRIESIAEVID  
AKTRYNSALIGVRIDAROFSSVPTSRSHLRRTIRVSPNDPATRTTGMDGTFKV  
AVSNRPAMFVFDI VILHKRCGLDGRVASAMVDWALYQIQCYCDELVPGRGQERFT  
CNCLQORNDAPVLQDLASFRGMAVYASNVAVADMPASAYLFFHGNVDRFT  
YAGSARRAKTVALVSNDPCDQYRAKVEHVDGDIARYSIQTEVTAFCGTSQSA  
ORWGMILTSLETEVTSFKYGLDAVAVMPGAIIEVDAPRAGRTNGRVSADGRT  
VVVDS PVVVRREDTLLINLPGTAORRTSRSDGSLTYSASQSPVAAEVAESA  
DUKTOLEFRVSVTESDGLTFDLSALEYNOSKRSALDNCTRIRAPRVSILPVSQAPT  
DVTLATYSVDQGIATITMMVAMQQAASVAATVEMRDNEMVTAAGTSGSLEVRN  
IYAGTYVARAINALDVALSPYASAEETLTKTSPPVVGTLLIATVFGIRIDMAP  
PTGPIADVERTLEMSKAPDRASATLGDFAFPANTHTLMAGATFEYMAIYVXKS  
NGAWYPSGNGVLSGSDASDILSYMKQIGKTQIAGLDSAVDSIDPPMAGSDEFT  
AGDGTVYGVSTQSVLEBANALAOQVTTMOATQAOSSAVOVAOQAAEBCGLAA  
MTTKTQIAPANRITLGIAGFGEVENDKVTESQVILADRPVILHPNKNVTTPVVO  
GGQVFMDSAFIADGTTTSKIDSIOSSNFWAGOTGMRLSKGLFENNNGNAGGRVD  
TSALTQIYDANGTLRVLGWM"  
/gene="RSC1699"  
3662.4318  
/note="synonym: RS02888"  
CDS  
3662.4318  
/gene="RSC1699"  
/function="miscellaneous; unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="HYPOTHEITICAL PROTEIN"  
/protein\_id="CAD15401.1"  
/db\_xref="GI:17428716"  
/db\_xref="SPTREMBL:Q8XY09"  
/translation="MPAGLTVYGEHGFVQITAEYGNLALRQKSTVTPDSCGSGSTTS  
ASRPVCIQSANPVALIGSSNGADWTFESAQAQPFVYVDEPAPASGQPGIQVF  
KADGTLAFDSGLAYLKVSGVTPPSGAPAYGQAMASPLTAGNVAACTSYTGTYAV  
PCTDTLFPVADHYTTANGKLTLYTHRGANNGDGSGLKIDTANPQIVLVDSQL  
"  
gene  
4387.4878  
/gene="RSC1700"  
/note="synonym: RS02889"  
CDS  
4387.4878  
/gene="RSC1700"  
/EC\_number="3.2.1.-"  
/function="elements of external origin; phage-related  
functions and prophages"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE BACTERIOPHAGE ENDOLYSIN PROTEIN"  
/protein\_id="CAD15402.1"  
/db\_xref="GI:17428717"  
/db\_xref="GOA:Q8XY08"  
/db\_xref="SPTREMBL:Q8XY08"  
/translation="MPYTDPSVYGGRNLAFLDMVGFSEGTNGHQPTREDRGYDLVG  
GGLFVSADHPRIELVLEPRLGIKSTAGAYQIILARYDPRQLRVLDVDPGAPAOAIA  
VOQIRERGLADIQGRLLAAIAKCKNIWASLPGAGYQHEQKFEALRAHYLLCGGGE  
VXG"  
gene  
4878.5186  
/gene="RSC1701"  
/note="synonym: RS02890"  
CDS  
4878.5186  
/gene="RSC1701"  
/function="miscellaneous; unknown"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Framed"

/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE LIPOPROTEIN TRANSMEMBRANE"  
/protein\_id="CAD15403.1"  
/db\_xref="GI:17428718"  
/db\_xref="SPTREMBL:Q8XY07"  
/translation="WTSEEMAVAAKSGSALSGVIALRPLDGTMMQRMLSFVSLGI  
GCLGAAVERFLVLPQSYTHMLAVASAAVFGLAIVNNAMQOIEILDLRRFLAKE  
"  
gene  
5189.5464  
/gene="RSC1702"  
/note="synonym: RS02891"  
CDS  
5189.5464  
/gene="RSC1702"  
/function="miscellaneous; unknown"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE TRANSMEMBRANE PROTEIN"  
/protein\_id="CAD15404.1"  
/db\_xref="GI:17428719"  
/db\_xref="SPTREMBL:Q8XY06"  
/translation="MLTISINQVANAILFVGSIMAVLTHQVPTFGALVIALVPAA  
LGNIVAHGTCHSMEVALNVAVALCVMAFWRLRVRFRFALGKGRG"  
5464.6009  
/gene="RSC1703"  
/note="synonym: RS02892"  
CDS  
5464.6009  
/gene="RSC1703"  
/EC\_number="3.4.-.-"  
/function="elements of external origin; phage-related  
functions and prophages"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE BACTERIOPHAGE-RELATED SIGNAL PEPTIDE  
PROTEIN"  
/protein\_id="CAD15405.1"  
/db\_xref="GI:17428720"  
/db\_xref="GOA:Q8XY05"  
/db\_xref="SPTREMBL:Q8XY05"  
/translation="WPGIKGTATIAAATMLGALVGAAGMTQAHRYDKDIATLQRD  
OAEKROQARDEAOEROKERKAGLEQJDELRHTEELNNKTERARLAALAG  
TROLSIRARCTAPADGVSTTSGAGLGGGAATAILYGEDNADIQALVADADAAAL  
RFQGRERAIQTRDTSVDARR"  
complement(6117..7304)  
/note="predicted by Homology"  
/insertion\_seq="ISR018 or RS05932"  
complement(6273..7238)  
/gene="RISR018"  
/note="synonyms: RS02893, RSC1704"  
complement(6273..7238)  
/gene="RISR018"  
/function="elements of external origin; transposon-related  
repeat\_region  
gene  
repeat\_region  
CDS  
Query Match  
Best Local Similarity 15.0%; Score 634.6; DB 1; Length 189050;  
Matches 1513; Conservative 0; Mismatches 1234; Indels 48; Gaps 9;  
QY 170 CCGCAGAAATCCATCGTCCGCCGCCCGCAGCGCAGCGCAGCGTCCGTTCTTCACAGACA 229  
DB 137989 CAGCCGACCCCATCGTGCATGCGCGCAGCACGCGCGCATCTGCTCCCGCTGACGAGC 137930  
QY 230 TCCAGAAATCCATCGTGCATGCGCGCAGCACGAGCGCTTACGAGTCCCGCAGCGGAT---CC 286

Dh 137929 TCGACGAGCTTACTGATTCGAGCCGACGGGCACTACCGCTGGGCGCGCTTCC 137870  
Qy 287 ACGCTATCGGAAATACGACTGTATCGATCTCGACGTGCCAGGCTGAGCCGCGCTTTC 346  
Dh 137869 ACGCTACTCGAGGCGCATCGCGCGACGCGCTGACCTCGGCTGCTTCGAAAGCTTCT 137810  
Qy 347 GGAAG---TCGTCCGCGGCACGACATGCTTCGGGCCCAACGCTGCCGCAATGATGC 403  
Dh 137809 GGAATGCCGTGTGCGCGCCACGCGCGCTGCGCTGTGTATCGACGAGCGGGGAGC 137750  
Qy 404 AGGTATGAGCTTAAATCGACCGCAATCGATGATG---ATTGCGCGGCTCG 460  
Dh 137749 AGCGATCTCGCGGGGTGCGGCTTACCGATCGGTCGCAAACTTGGCGCGGCA 137690  
Qy 461 ACCGAGACACGAGGACGAGGCTGTGTGTTGCGATGCGATGCGACATCT 520  
Dh 137669 CCGCGAGGCGCTTGGCGCGCACTGCGACATGCGCGCAATGTGCGACAGGTGC 137630  
Qy 521 ATGACACGAGCGCGCTCGCTTATCGATGTCGCGCTTTCGCTGAGACGAGCGCAAA 580  
Dh 137629 TCGATGCCGCGAATGCGCGCTGTTCGACGTGCGCGCACGACCTGCGGGCGGGCA 137570  
Qy 581 CCGCTCTGTGCTCATGATCGATCTCATTAAGTTGACCTTACGAGCGCTGTCATCT 640  
Dh 137569 CGCGCTGACATCGGATCGACATGATCAAGACGCGCACGCGCGCGCATCT 137510  
Qy 641 TCAGGA-----CTGGCTCAGCTTCTACGAAATCCGAGACTCTCTCCCTG 688  
Dh 137509 GGGACGACTCGCGCGCTGTACCGCGCGCGCGGACCTCGAACGCGCGCGCTGCGC 137450  
Qy 689 TCCTGAGCTCTCGTACCGCGCATTTATGTACTCGCTGAGTCTGCGAAGATC----- 743  
Dh 137449 CATTCGATCAGCTTCCGCGCATCGTGTGGGAAATGATGACGCGAAGCGCGC 137390  
Qy 744 -TGAGCGCATCAAGATGATGATTACTGAGAGCGCGCATGCGCATCTCCACTC 802  
Dh 137389 GCGCGCGCGGATCGGCAAGGCTTACGTGCGCGAGTCCAGACCTGCGCCCG 137330  
Qy 803 CGCGGAGCTTCGATGAGGCGCATTCATACCTGAGAGATCCGCTCCGCGCA 862  
Dh 137329 CGCGCAATCGCTTGGCGCGCGCGCGCGCTGCGACCGCGCGCGCGCGGATTCGCGCG 137270  
Qy 863 CGAGCATGCTGCGCTCGGACTCTGAGGTGATTAAGCGCGGTTCGCGGAGCGC 922  
Dh 137269 GCGACATGCTGAGCGCGCGCGCATGCGCAAGCTCGCGACCGCGCGCGCGATGG 137210  
Qy 923 GAGTGAATCCGACGCGGCTCATCTGCTGATTTTCGAGGTGATCGGCGCTGAGCG 982  
Dh 137209 GCTGACCGCGGCTTCTGCTGATCGCGGTGTTCGAAAGTCTGTGCGCTGAGCA 137150  
Qy 983 CGAGCGCGCGTTTACGCTCAACATAGCTTTCACCGGCTCCCGCTCATCGCGCG 1042  
Dh 137149 CGAGCGCGGTTCACGCTCAACCTCACCTTCGACCGCTTTCGCGGACGCGCA 137090  
Qy 1043 TGAAGATATACGCGGACTTACGATGATGATCTCTGGAACATCGACCACTGCG 1102  
Dh 137089 TGCGCGCTGCTGCGGATTTACCGCGGTGACGCTGCTGCGTGAATGCGCGCGC 137030  
Qy 1103 ACAAGCTTGAAACAGCGGCTTAAGCTTTCAGAGACGCTGAGAGCGATGATC 1162  
Dh 137029 CCGTCCCTTGGCGCGCGCGCGCGCTGCAAGCGGCGGTGTGAGACCTGCGAGC 136970  
Qy 1163 ACTGAGATGAGCGGTATCGAGGTGCGAGCGCGCGCGCGCTGAGGATTCAC 1222  
Dh 136969 ACCGCGCTTACGCGCGGTGACGTGTGCGGAATGAAACCGCGCGCGACGCGCG 136910  
Qy 1223 GAGGCGATGTTTCCCGTGTGCTCAGAGCGCG-----TTAACAGCAAGCTTTC 1276  
Dh 136909 ATGGGTGTCCATGCGGTGTGTTCACACGCGAGCTGCGATGAGCAACCCACAGG 136850  
Qy 1277 GTGTCACTCTGTGAGAGGCTCGGAATCGGCTGACACGACGAGACTCTCAGC 1336

Dh 136849 GCGCGCGCGCAAGCGTGTGGGACCGTGGCTACGCGATGACGACCCCGCAG 136790  
Qy 1337 TGCTGTGATCATCAGCTTCTACGAGACGATGAGGAGACTGTCTCGCTGAGACATCG 1336  
Dh 136789 TGTGGCTGAGACCAAGAGCTTCCAGCTCGACGCGCGGTATCTAACACTGGATGGG 136730  
Qy 1397 TCGACGAGTGTTCGCGCGGACCTTTCGACGACATGCTGAAAGCTGCTGTTTC 1456  
Dh 136729 TCGATGCGCTGTTCACCGCGCGCTGCTGATGCTCATGTTGACGCTTCAACCGGATGC 136670  
Qy 1457 TCGCGCGCTTAC-----TGAGAACATGAGGATGAAAGAGTGGCTGCTTCGCG 1510  
Dh 136669 TAGAACCTTCGCGCGCGACCGCATGCTGTGACGCGCTCCGCTGTGCGC 136610  
Qy 1511 CTGCCAGCTGAGACCGCGCGCGAGCAAAAGCGACCAACGCGCTGTGAGACGATA 1570  
Dh 136609 AAGCTCAGCGAGGTGCGCGCGCTGCTCAATGCGACGCGCGCTGCTGCGCGCT 136550  
Qy 1571 CGTGCACGCGCTTTCGCGCGCGGTGAGACGCTGCCATGACGCTGCGGTGT 1630  
Dh 136549 GCTTCGACGAGCTGTTCTTTCGACGAGGAGCGCGCGACGCGCGCGCTGATTCG 136490  
Qy 1631 CGCGCGCGAAGCGCTACGATCGAAGAGCTTTCGCGCGCTTTCGCGCGATTCGCGCGC 1690  
Dh 136489 CCAAGACCTTCACTGACCTTACGACCACTGCGCGCTGCTTCGCGCGCTTCCGCAAG 136430  
Qy 1691 GCGTCCGCGACGAGGCGCGCGCGCAACATTTGTCGCGGTGTGATGAGAAAGCT 1750  
Dh 136429 GCTTCGCGACGCGCGCGGTGAGCGCGGTGACCGCGGTGCGGTGTGATGCGCAAGGCG 136370  
Qy 1751 GGGACGAGTGTGCGCGCTTTCGCGGTGCTGAGTCAAGCGCGCGCTTACGTCGATCG 1810  
Dh 136369 CGAGGAGGTGCGCGCTGCTGCGGATTCAGCGCGCGCGCGCGCTGATGCGCGGTG 136310  
Qy 1811 ATGCGGACATCGCGCGGAGGTATTCATCACTCTCTGATCATGTGTAGGTAAAGCTG 1870  
Dh 136309 ATGCGGACCGCGCGCGCGCTGCGACGCGCATCTGCAAGGACGCGCATCGCGGTG 136250  
Qy 1871 TGTGACGCGACCATGCTGATGAGCAACTGTATGCGCGCGGATTCAGCGCGGTG 1930  
Dh 136249 GCTGACGCGAGCGCATGCTGCGCGCATGCTGCGCGCGGTGACCGAGCTTTCGACGTGC 136190  
Qy 1931 TCGTGAAGAGCGCGCGCTGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1984  
Dh 136189 GGTGATGATGCTTCGCGAGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 136130  
Qy 1985 TTCAAGACCTTTCGATGCTGCGGTATGATCTATCACTGCGGATTCACAGGCTTTC 2044  
Dh 136129 CGCGCGATCG 136070  
Qy 2045 AGGCGGTGATGATGATCATCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104  
Dh 136069 AAGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 136010  
Qy 2105 TCGAAATAGGCGCGGAGACAGGCTGCTGCGCGCTTCTCGCTGAGCTTTCGATCTTCG 2164  
Dh 136009 TCGGATATACCGCGCGGAGTGTGTGCTTGGCGCTGCGCGCTTTCGACCTGTCG 135950  
Qy 2165 TCTATGATGTGTTGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2224  
Dh 135949 TCTATGATGTGTTGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 135890  
Qy 2225 AGCTGCGGATTCG 2284  
Dh 135889 GCGTGCAGATCG 135830  
Qy 2285 ACTGCTGCGCGCGCTGATGCGGATGCTGTCGAGCAATTTTGAAGGTGCGCGCGATTC 2344  
Dh 135829 ACTGCTGCGCGCGCTGCTGAACTGCTGCTGAGCGAGCGCGCGCGCGCGCGCGCG 135770  
Qy 2345 TCGCTAGGTCTTTCGCGCTTTCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2404  
Dh 135769 TGCGCGGATTCG 135713

```
QY 2405 GCGAGCTCCAGGCCATCAGGCCCGCGCTGCTCGGTATCAAGCTCGGCGGCGCCACCGAAG 2464
Db 135712 CCCGCTGCGCGCCCTCGCGCGCGCGCGCTGATTCGCAATGGCGCGCGCCACCGAAG 135653
QY 2465 CGTCGATCTGTCATCGGGTACCCCGTAGGAAACGATCGACCTATCGTGGCGAGCATCC 2524
Db 135652 CGTCGATCTGTCATCGGGTACCCCGTAGGAAACGATCGACCTATCGTGGCGAGCATCC 135593
QY 2525 CCTACGGCGCGTCCGTCGCGCAACAGACGTTCCACGTCGTCGATGAGGCGGCTCGAAACCGC 2584
Db 135592 CCTACGGCGTACCGCGCTCGGCAACAGCGCTACCGGTCGTCGACACACACTGCGCGACT 135533
QY 2585 GCCCGGCTGCGGTTCCGGGCGCACTTACATTGGCGGGGTGCGGCTGCGACCTGGCTACT 2644
Db 135532 GCCCGGACACGATCGACCGCGGATCTCTGATCGGCGGCACTGCGCTCGGCTGAG 135473
QY 2645 GCGCGGATGAGGAAGAACGCGCGAAGAGCTCTGTCGTCGACCCCGAAGACCGGAGAGCGCC 2704
Db 135472 AGAACGATCCGAGCGCGACGCGCGCGAGCTTCTTCCACATCCGCTCGACGCGGAGCGCC 135413
QY 2705 TCTAACAGACCGCGCGATCTGCGCGCGCTACCTGCGCGGATGAAACATCGAGTTCTAGGGGC 2764
Db 135412 TCTAACCGACCGCGCGATCTGCGCGCGCTACCTGCGCGGATGAAACATCGAGTTCTAGGGGC 135353
QY 2765 GTGAGGACACCAATCAAGCTTCCGCGATTCGCGCTTGAAGCTCGGGGAAATCGAGGAA 2824
Db 135352 GCGCGGACCTTCCAGGTCGAGATCGCGCGCGGATTCGCGCGGATTCGAGGAAATCGAGGCG 135293
QY 2825 CGCTCAAGTCGATCGGAAAGTACGCGCGGATTCGCGCGGATTCGCGGAAAGCGAGCGCG 2884
Db 135292 CCTGCTGCGCGACGCGAGGCTGCGCGGATTCGCGCGGATTCGAGGCGCGCGCGCGCG 135233
QY 2885 CGAACAGCTCTTCTAGGCTATGTGTGTCGCGAG 2919
Db 135232 GCACACAGCGCTGCGCGGATTCGCGCGGAG 135198

RESULT 13
AF516145 40156 bp DNA linear BCT 03-OCT-2002
LOCUS Lynbha majuscula barbamide biosynthesis gene cluster, complete
DEFINITION
sequence.
ACCESSION AF516145 GI:23452290
VERSION AF516145.1
KEYWORDS
SOURCE Lynbha majuscula
ORGANISM Bacteria; Cyanobacteria; Oscillatoriales; Lynbha.
REFERENCE 1 (bases 1 to 40156)
AUTHORS Chang,Z., Platt,P., Gerwick,W.H., Nguyen,V.A., Willis,C.L. and Sherman,D.H.
TITLE The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroethyl starter unit
JOURNAL Gene 296 (1-2), 235-247 (2002)
AUTHORS Chang,Z., Sherman,D.H. and Gerwick,W.H.
TITLE Direct Substitution
JOURNAL Submitted (28-MAY-2002) Microbiology, University of Minnesota, 421 Delaware Street, SE, Minneapolis, MN 55455, USA

FEATURES
source
1..40156
/organism="Lynbha majuscula"
/mol_type="genomic DNA"
/db_xref="taxon:158786"
5531..5818
/gene="barA"
5531..5818
/gene="barA"
5531..5818
/function="probably involved in the activation of leucine to initiate barbamide biosynthesis"
```

```
/note="peptidyl carrier protein"
/codon_start=1
/transl_table=1
/product="BarA"
/protein_id="AAN32974.1"
/db_xref="GI:23452291"
/translation="MSNSTELPIDVVVPOTETEKOLAEIKWLVSTISIEPRFMD
IGNSINDYEVNQVTEKMGVSIKARFPDKSTIAELSEIDAVNEQTH"
5850..6752
/gene="barB1"
5850..6752
/gene="barB1"
/function="probably related to chlorination"
/note="similar to the tryptophan StryB2 protein and the coronamic acid Cmb protein"
/codon_start=1
/transl_table=1
/product="BarB1"
/protein_id="AAN32975.1"
/db_xref="GI:23452292"
/translation="WKTAKAELSYVLSVEKQFPDNGYIGPTLPOPEITLAME
IRMDLDRETAAPFNSKMYDRHLDIALSDIISHPKIAHVSLLGSNVLCMTAMF
PKYDDEGTDMHQAESFVEFGSKLVPESHGDMELTAWAFSEATRENGCMKMP
GTHNTWYDEDRNIEFBDKINOKLTGKGTGYDYDKLDPNMEPDESOAVHLE
MERQFILTSCRHMGSEPNSSIRGMSRTRVPDVRVYPMDSFHFGEKMLDR
RYATVYAGEPTYKINKIRPLQ"
6812..7705
/gene="barB2"
6812..7705
/gene="barB2"
/function="probably related to chlorination"
/note="similar to BarB1, tryptophan StryB2 protein and the coronamic acid Cmb protein"
/codon_start=1
/transl_table=1
/product="BarB2"
/protein_id="AAN32976.1"
/db_xref="GI:23452293"
/translation="WKTITAEPRYLYSVEKQFPDNGYIGPTLPOPEITLAME
IRMDLDRETAAPFNSKMYDRHLDI KALNDIISHPKIAHVSLLGSNVLCMTAMF
PKYDDEGTDMHQAESFVEFGSKLVPESHGDMELTAWAFSEATRENGCMKMP
GTHNTWYDEDRNIEFBDKINOKLTGKGTGYDYDKLDPNMEPDESOAVHLE
MERQFILTSCRHMGSEPNSSIRGMSRTRVPDVRVYPMDSFHFGEKMLDR
YSTVYAGEPTYKINKIRPL"
7839..8665
/gene="barC"
7839..8665
/gene="barC"
/function="proofreading or editing role or chlorination role"
/note="thioesterase II"
/codon_start=1
/transl_table=1
/product="BarC"
/protein_id="AAN32977.1"
/db_xref="GI:23452294"
/translation="WQTDSSWLPFDETPKCEASKMLFCHYAGAGSITQDMSRAL
PDEVELVAIQLPGRKRCYEPMSNTEINIVLQELSPLSKPAFEGHMGALIAF
DLIRLRQOELGPOFLWNGFDPALPKSKSMWQLPDEFAQALIDYNGIPIVIA
ONEELFELLRLRSDIKLGEYIKOBEPLDCRIIALYSTEDSFVSKGMYQEQT
KGEPLVTFPGDFLHSEELVNTICTELQRYIKANDNCSTKVALQKS"
8700..10253
/gene="barD"
8700..10253
/gene="barD"
/function="adenylation; involved in activation of leucine"
/note="sequence similar to adenylation domains of NRPS"
/codon_start=1
/transl_table=1
/product="BarD"
/protein_id="AAN32978.1"
/db_xref="GI:23452295"
/translation="MGKVIANLVTRHSKSPHKCAITDGTLESYGVWEASNPARK
```



LIEMGCKGDRVNVLASKQITLIMVGLIGFACGICIVHEDLDMKPMVNRINKITITETIPS  
VITLTDQELHIVKSENVPOEKIVILEQALPITSTVEKRLPQISEBIDIAACYTGS  
GSTGDPKVLIVKSENVDPFEGTREVDPEDSRGVSPHSPDLSVIDIDIPYOGA  
MLYISDVVHLEHLEFVIONKSNVTHFAFGSMGLIADQEKEDVILKIDLTITLGGAD  
VPDIKVIQKMLKGENIKITINGTPTBETCCACATITIDRPNRRELYPTGKPLKNAK  
LYVDQOQEIYDEPIPEBLIGGTQVAGVYNNRQSKERLVEKGTCTKTDICCK  
VLADGNLEFVRNDNEVNGVRYVHAINIKOVINRPHVNSVUTLESKGEKMLAA  
ALMLANETSKDASAKQIKIENHOBELPQYVMEPRQIVYMEKFPQLSGKTRKLTISILE  
RDLASKVH"  
10313..15433  
/gene="bars"  
10313..15433  
/gene="bars"  
/function="involved in activation of tricholeucine and  
condensation with malonyl group"  
/note="NRPes/PKS; A, PCP, KS and AT domains are located on  
Bar"

gene  
CDS

/codon\_start=1  
/transl\_table=11  
/product="Bar"

/protein\_id="AAN32979.1"  
/db\_xref="GI:23452296"

/translation="MKNDPMGCIHQKNDVFNHVKSTSGSLNIPNSPTLLCLLPR  
AALTKNGITVQADGSEQYQYQDILLDAKRIAGLRQMDLQPDQVILLQEBHDFI  
PAFWGCLGGRIPVPLSIAPYQOLNNTKILHNAQMQLHLEPLILTNRELAPRISLS  
THNLIDPQGTVIDLKRQDPTQOIRGQPELILLLTSGSTGPKGMVSHRNILS  
ESAGTIQKMHFDEAVTLNMMPLDHVGGIIMLVNVAVLGGQVHAIPKYLIONLHM  
LILBHOATISMGPNFAYSLINDLVQDAGSMDLSISLVNGSEISKVARRL  
KLMRQGLAETVLPARFGMSTSGITTDPSLSNTSDMFPGLGPIYAGACRLV  
NDNDEIVHEGAIIGLQVKGASVTSGYONPEINQAFADMGFRGDIYGLACRLV  
VGRKDVLLINGVYPAHEIAVEEDVGVISFTAAICSAIGSSISDOLAFSSA  
VEANIEITPEVLKGLIKKIRGAVVONIGINDYLLPVKESIKPTAKISIKROKLY  
FEAGEFPLIETVGGSLRIMGSOETALALPOTVEQVAVNIWODVGLGDKYGOENF  
ELGHSLLILTVGOSLRSFGKELSIAMFHPTEALQAPYTOPDPSDASAOOCYAR  
AQIOEVRKADSTELAVIAGMGRFPAGDNLQFQWNLONGESTIFFSDSLASGID  
PKLSNPYVAVASPLSDVEYFDAHFEYTRAEFMDPQRLMECMESLEADGAD  
PLTYSGSLVAGAMLSYFIHMLPNRDTDEHDLKVLTPDSMGGLQVILANDKDY  
LITRISHKLANTPGSVINIOTACSTSLVTVAVOSLNGECDLAAGSVQVORQV  
HLHQBGLIASPDGCRAPDAQEGTIPNGVGLVAKRLERALODGHIYAVITGSAL  
INDGSKVAYTAPNGEQTIVAAEIANINPEITIVYEAHGTITLGDPIELGALS  
QAFQKTRQKQFCALGSKVIVNGHQLITSGVGLIKVIALYHKQIPSLHFOPTNEK  
IDFANSPFVNTMLSEWRNGHPRRANGHFGGTNAHILLEBPNAQKOTNEEDG  
AKLEDRSHTLTLAKTAPALQALVKRQRLSDNTELELDICTACVAGAEHRL  
AIVSDSIPDLRCKLANVLDNAELSBETOELAQOCTASVRAQIALFNTGQGOYLN  
MGRLOYTOSVFRDAIDCCQDILQSELESLIDLITPOTANAAARNNTAYOPL  
PAIEYLAQVMSWGCIQPAVYMGHSVEGVAAATAGVDTGTAKLIARRLMQQLP  
AGXANVMSAEERVSLSIPHEKISIAIANGPSVVISGVAELGAIISQLESAGV  
KTKSLQVSHAFHSPLMEPIIAEFVVAQILYHSPRIPLISNVGVKADQIATAQV  
VNHILROPVRFHSHMBALSEFGYDCLFIEGPRPILLMGRQCLPQGTGWLPLATPV  
DEQOMLSIAGLYQGVAVNMSLNOEYSYQKTLPTYPFRKMYIKIKEYTNEOKFT  
SATSNOKNNTFIRLLNGNVLDILOYLKTDLGREEVKLSKPAQILINLVHQBNE  
LITKDKERKIDLMQMLQMDSENLKLSLEAKFSQNS"

gene

CDS

15476..16990  
/gene="bars"  
15476..16990  
/gene="bars"  
/function="involved in addition of O methyl group"  
/note="contains methyltransferase (MT) and ACP domains"  
/codon\_start=1  
/transl\_table=11  
/product="Bar"

/protein\_id="AAN32980.1"  
/db\_xref="GI:23452297"

/translation="MSDDLKMMNNUOERLSTLSPEOROFPEKTIANKPHOQSGOF  
IAEKSTBERLSRYRALDHSNTADYARFSEVDPFMEVLNIDPHDPQOTSL  
SLQAQAEKVALFRGIDFSSIKYMDIGCGSHLIDATLVNHLQLDGNYISPEYK  
AGEOKIOGLISYDRILYLRNDASAKQPLDIDTLIFSCLVNIHILKKEVEVFNIGOHAN  
DSGEFVAEISNPLPIRIDAKSTAYVTRSKMAQOLLNKRKEVEVFNIGOHAN  
LYDNFNTONLARLCOGYDENAKDHVGEHLLKRLAYLILTYOQSSFTKERT  
LRNDEKTONLTPYSKIENSNGKPIVLPFRBEGSTPQCGTRDPDROQPLSDLS  
DAVOGLAFLKTVAAIILKETQAEINETSITVGVDSVLSLEJNRNQAOLGVDISV  
ADLMGDNISRLAIKVSLLHRNAQGAQDLSEKMDGMLQELIADKKMRBGRLL"  
16987..25650  
gene

/gene="bars"  
16987..25650  
/gene="bars"

CDS

/function="responsible for the recognition and activation  
of phenylalanine and cysteine and for the release of the  
barbamide intermediate from the assembly line"  
/note="NRPes; two NRP modules are located on the protein  
containing C, A, MT, PCP, Cy, A, PCP and Te domains"  
/codon\_start=1  
/transl\_table=11  
/product="Bar"

/protein\_id="AAN32981.1"  
/db\_xref="GI:23452298"

/translation="MNLSEFLQELVLSGMQFAEEGVCFQAPPADSTQVLAOLKOH  
KRDLITLOEHPVLOVPLIGQGGIIMPLQMLPPDNNVNSVATRIYSOVNNTTQO  
OTPEAKRHRHLSTPRKCEPTIRHOSBEDPVOIDASWDSNELOTOVAAHRHP  
FDLOTDPVWRVMTRESQEBHILTLIHILAMDSANITVLEISELOAHACAGVAVD  
LPSLOHTYQDVYKQQQLVEGSKESLMTYQQQLAGELPVNLPTDRPPIQTNGR  
AVYRQPLPHELVTVQKALQABEATLVNLLAOLVILHRYTGQEDLVGSPSTGRTR  
PEFTSVGVFVDSWVRKAVGSLSPRELFQVOTVYDALADODYPSLVEKLOPE  
RDLRSPLIPQVPELHNPLOSETQQLFLGETKTLVHMGMEVFTEFLPQYSELDVL

Query Match 14.4%; Score 609.4; DB 1; Length 40156;  
Best Local Similarity 52.9%; Pred. No. 2e-86;  
Matches 1499; Conservative 0; Mismatches 1291; Indels 45; Gaps 7;

QY	179	CCATCGTCCGCGCCGCGAGGCGGACGCTCGGTTCTCTGACAGCATCCAGAT	238
DB	21581	CAATGTGCTCGCGCGAGGCGGCTTTCACACCTTTCCTCACAGATTCACAAAG	21640
QY	239	CCCTACTGCGCGCGCGGACAGAGCGTTTACGCTCCCGAGCGGATC---CACGCTTATC	295
DB	21641	CTTATGTGCTCGCGCGGACGATTTGACCTTGGGTAATATTAACAGCATGTTATC	21700
QY	296	GCGAATACGACTGTAAGGATTCGACGTCGCGAGCGTGAAGCGGCTTTCGAAAGTCG	355
DB	21701	TGGAACTGCGACTGTGAGATTTAGCTTTAGACCGGTTAAGTCAGGCTTGGCAGCAATGA	21760
QY	356	TGCGCGCGCGGACGATGCTTCGCGGCGCACACGCTGCCGACATGATGAGGTATCGAGC	415
DB	21761	TAGACCATGACGATGCTGGAATGTGATTTTACCAATGTTGAAAGCAAGCTCTTG	21820
QY	416	CTAAGTCGACGCGC---GACATCGAGATCATGATTCGCGGCGCTCGACCGGACACAC	472
DB	21821	AACAGCTTACCGCTTACGATTTGAATTTAGCTTCGCGGCTCGACCAACAAACAAATCG	21880
QY	473	GCGAAGCGAGGCTCGTGTCTGTCGGAATGCGATGTCGACCGCATTTAGACACCGAGC	532
DB	21881	TCTCAACCGAATCAAGAAACCATTTGTTATCGCTGTCACGAGATGTTCCCGCTGGAG	21940
QY	533	GCCCTCGCTCTACGAGTGTGCGCTGCGCGTGGCGAGACGGGCAACCGGTCGTCG	592
DB	21941	AGTGCCCATCTTAAAGATTCGTGTACTGCTTACCGAGATCAACGCTTACACT	22000
QY	593	TCAGTATCGATCTGATTAACGTTAGCTAGGACGCTGCTCCATCATTTCAAGACTGCG	652
DB	22001	GGACCTTGANCGCTTATGCTATGCTATGCTGACAGATGATGATGTTGGCAACAATGCG	22060
QY	653	TCAGCTTTCAGAGATCCGAGACTCTCTCTGCTGTCGAGACTCTGTAACGCGATT	712
DB	22061	TGCAGCTATACCAAGAACCCAGACAGTTTCTTCAACCAAAATTAGACTCATTTGCGGACT	22120
QY	713	ATGTAATCGCGCTGAGCTCTCGCAAGAGTGTAGCGGCAATCAACGATTCGATTAAT	772
DB	22121	ATGTTTAGAGAACTAGTCTCAAGACAGCGCCCAATATTCAGCGCTTCAACAGATTT	22180
QY	773	GGAAGCGCGGATGCGAGCTCCCACTCGCGAGAGCTTCCATGAAGCGCATCAT	832
DB	22181	GGTGGAATCGTTAGAGACACTCCCCCTGACCAAGAACTTCTTGTGTAAGCAACTG	22240
QY	833	CTACCTCGAAGAGATCCGCTTCCGCGACACGAGCAATGCTGCTCGGATCTCTGGG	892
DB	22241	CCACACTAGACCACTGAATTTAATTGTTATCGGGCTGAAGCTGCTGCCCCAGATGGC	22300

QY 893 GTTCATTTGAAGCGGCGTGTCTGGGAGACGGGGCTGACCCCGACGGGCGCTCATCTGGCTG 952  
Db 22301 AGCAATTTGGAACCCAGAGGGAACAAAGCAGATTGACACCGTCAAGGGGCTCTGTGAGCCG 22360  
QY 953 CATTTTCCAGAGTATCGGCGCGCTGGAGCGAGCCCGGTTTACGTCAACATTAAGC 1012  
Db 22361 CGTTTCCGATTTCTTAAGCTATTTGAGAGCAAAAGTCCCAATTACAGATCAATTCACCC 22420  
QY 1013 TTTTCAACCGGCTCCCGCTTCATCGCGCGTGAACGATATCACCGGGAATTACGTGCA 1072  
Db 22421 TCTTTAACCGGCTACCTTTGCAATCCCAAGTCAATATTTAGTGGAGACTTCACCTTCCC 22480  
QY 1073 TGTCTCTCTGACATTCGACACCACTCGGAGCAAGAGCTTCGACAGCGCGCTAAAGCCTA 1132  
Db 22481 TACACTCTTTAGAGTCAATCAAAAAAGCAGACTCCCTTGTCTCAGCGTGTCAACGCTC 22540  
QY 1133 TTCAAGAGCAGCTGTGGAGAGCGATGATCACTGCGACGTAAAGCGGTATCGAGTCCAGC 1192  
Db 22541 TCCAGGGAACAATTGGCAAGACTAGACACCGCTATGTGGGGGCGTTGAGGTCCAGC 22600  
QY 1193 GAGA---GGCGCCCGGCTCTGGGGAATCCAGAGCGCATTTGTCCTCCGCTGCTC 1248  
Db 22601 GAGACTTTCGCGTCAAGCGCGGTATATCAACCCATGGGGGTTGTATTTACAGATACGC 22660  
QY 1249 AGGAGCGGCTTAACAGAGAGTCTGTGGTTCACCTCGTTGAGAG----- 1295  
Db 22661 TGGCGTGAATAGTGGAGAGAGAGAGATGGCTGCTCAATGATGAGCAGCTTGCCAT 22720  
QY 1296 -----GCTCGAATCCGCGGTGATACAGACAGCAGACCTCTCAGCTGCTGGATC 1348  
Db 22721 TGGACCAATTTGGAGAACTGTTTATATGGTCTTAATAACCTCTAAGTGTGGCTAGACA 22780  
QY 1349 ATCAGCTTAACGACGATGGGAGCTGTCTCTCGGTGGACATCTGACGAGATGT 1408  
Db 22781 ATTCTGTCTGAACAAATATGGTGTCTTACTCTTAATCTGAAATGATGAGAAAGCTCT 22840  
QY 1409 TCCCGCCGACCTTTCGAGCGACATGCTGAGAGCGATCGTCTTTCCTCCGCGGCT-- 1466  
Db 22841 TCCGAGAGGGTCTCAATGATGATGATCAAGACTACATCAATGATGAGCAACACTCG 22900  
QY 1467 ----CACTAGGAAACCATGGGGGTGAACAGGTGGCTGTTCGCTTCCGCTGCCAGCTAG 1522  
Db 22901 CAACCTCGATGTGGGCTTGGGGCGACATTTGCCCTCAACTACGCCCCCTCTCAACTTA 22960  
QY 1523 AAGCGCGGCGAGCGCAACCGCAACGCGCTGCTGAGCGAGCATACGCTGACGCGC 1582  
Db 22961 CCCAAGCGCTCAAGTTAAATGAACCTATGCCCTGTCTTCTGAAAGAGACCTTCATATTT 23020  
QY 1583 TGTTCGGGGGGGGGTCGAGCGAGCTGCCATGAGCTCGCGGTGGTGGGGCGGAGAA 1642  
Db 23021 TGTTCGTAAGCAAGTTGACGAGCGACAGAGCGATCGCCCTCATACACCCGCAAGCA 23080  
QY 1643 CGCTACGTACGAAAGACTTTCGCGCGTTCGCGGCGACTTGGCGCGCGGCTGCGGAGC 1702  
Db 23081 CCTCACTACATGAACTCTACACGGAAGCTCAGCGGCTCGGCGAGCAAGTTCAACAT 23140  
QY 1703 AAGGGGACGCGCGAACAATGTGTGCGGTGTATGAGAAAGGCTGGAGCAGGTTG 1762  
Db 23141 TGGAGACCAACCCCAACACCTATGAGCGTGTCTCATGGAAGAAAGCTGGAGCAATATG 23200  
QY 1763 TGGCGGTTCTCGCGGCTCGAGTCAAGGGCGGCGCTACGAGCGGATGATGAGCACTTAC 1822  
Db 23201 TGTCTCTTAAGCAATTTGATGTGTGTGACACCTTATCTCCCATTTGATGAGCGCCCTG 23260  
QY 1823 CGCGGAGGATATCACTACTCTCTGATCATGTAGGTAAAGCTGTGCTGACGACG 1882  
Db 23261 CCCAGAGGCTCATATGTCTTGTCTGAGAGAGAAAGTCAAACTGTGTGTGACCAAG 23320  
QY 1883 CATGCTGATGCAAACTGTATGCGCGCGGGAATCCAGCGGCTGTGTGAGCGAG 1942  
Db 23321 CTGCGCTAAACGTACTTATGATGCTCAGAACCACTGCACTGCTGTGTGT-----TG 23374

QY 1943 CGGCGTGAAGAGCGAGCGGCGACCAAGCTCCGATGATGCCATTCAGACCTTCGATC 2002  
Db 23375 CTTCACCAACAGAGAGATCATAGACATCCCCCTGAGGCAAACTTTAGCAGACGAT 23434  
QY 2003 TCGGTATGTCACTTACACCTTCGGATCCACAGGTTGCCCAAGGGGTGATGATCATC 2062  
Db 23435 TAGCTATGTATCTTTACTCTCGGTTCTACAGAAACCCCAAGGGGTGATGATCATC 23494  
QY 2063 ATCGGGTCCGCTCAACACATCTCTGACATCAACAGACGCTTCGAAATATAGGCCCGAG 2122  
Db 23495 ACCGAGAGCGCTTAACACATCCATCAAGATATTTAACCAAGGTTTACGTTACGCTTAC 23554  
QY 2123 ACAGGATGTGCGCTCTCTCTGATGATTCATCTCTGATGATGATGATGATGATGATGATG 2182  
Db 23555 ACCGATGTGCGGCTCTCTCTGATGATTCATCTCTGATGATGATGATGATGATGATGATG 23614  
QY 2183 TCTTGGCGGCGGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2242  
Db 23615 TCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23674  
QY 2243 ATTTGGGAGATTTGATTCAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATG 2302  
Db 23675 ATTTGGGATGATTAATGACAGCCCATCAAGTACCTCTGGAACAATGATGATGATGATG 23734  
QY 2303 TCGGATGCTGTCGAGCATTTTGAAGGTGCGCCGATTCGCTCGCTAGATGATGATGATGATG 2362  
Db 23735 TCGAATGTGATGAAATATCTGTGTGAGACACCCAGACCAAGTATGATGATGATGATGATG 23794  
QY 2363 TTTTCTGTGAGCGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2422  
Db 23795 TTTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 23854  
QY 2423 GGGCGGCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2482  
Db 23855 GGGCGGCGGAG 23914  
QY 2483 GGTACCCGCTGAG 2542  
Db 23915 ACTATTCATCTACCTGATGAGCGGAGATGAGAGAGATGATGATGATGATGATGATGATG 23974  
QY 2543 GCACACAGAGCTTCAAGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2602  
Db 23975 TGAATCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 24034  
QY 2603 GCGAATCTATCTTGGCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 2662  
Db 24035 GTACCTCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 24094  
QY 2663 CGCGAAGAGCTTCTGTCGACACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2722  
Db 24095 CGAATGCAAGCTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 24154  
QY 2723 TGGGCGCTACCTGCGGATGAGAAATGAGATTCATGAGGCGTGAAGCAACCAATCA 2782  
Db 24155 TGGGCGCTACCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 24214  
QY 2783 AGCTTTCGCGAGATGAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2842  
Db 24215 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 24274  
QY 2843 AGCTACGAGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2902  
Db 24275 CAGTTAAGAAACG---TTGT 24331  
QY 2903 CCTATGTGTTCGAGAGGAG 2962  
Db 24332 CCTATGTGTCTTCACTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 24391  
QY 2963 CCGAGCGATGAGC 2977  
Db 24392 ACATGCGATGAGC 24406

RESULT 14	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL
SC0939132	302007 bp	DNA	linear	BCT 11-FEB-2003													
Streptomyces coelicolor A3(2)	complete genome; segment 29/29.																
AL391313	AL355755	AL391338	AL391395	AL391388	AL391454	AL446003											
AL501465	AL450222	AL450289	AL450350	AL450432	AL450433	AL512944											
AL596144	AL645882																
AL391313.1	GI:24413926																
Streptomyces coelicolor A3(2)																	
Streptomyces coelicolor A3(2)																	
Bacteria; Actinobacteriales; Streptomycetales;																	
Streptomycetaceae; Streptomyces.																	
1																	
Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L.,																	
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kleiser, H.,																	
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,																	
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,																	
Huang, C.H., Kleiser, T., Larle, L., Murphy, L., Oliver, K., O'Neill, S.,																	
Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S.,																	
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,																	
Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrett, B.G.,																	
Parkhill, J. and Hopwood, D.A.																	
Complete genome sequence of the model actinomycete Streptomyces																	
coelicolor A3(2)																	
Nature 417 (6885), 141-147 (2002)																	
21996410																	
12000953																	
2 (bases 1 to 302007)																	
Bentley, S.D.																	
Direct Submission																	
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces																	
Sequencing team, Sanger Institute, Wellcome Trust Genome Campus,																	
Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk																	
On or before Oct 26, 2002 this sequence version replaced																	
gi:20520855, gi:9843813, gi:9857338, gi:9857140, gi:20520741,																	
gi:20520786, gi:20520767, gi:20520847, gi:20520681, gi:20520707,																	
gi:20520788, gi:20520682, gi:20520861, gi:20520876.																	
Location/Qualifiers																	
1. .302007																	
/organism="Streptomyces coelicolor A3(2)"																	
/mol_type="Genomic DNA"																	
/strain="A3(2)"																	
/db_xref="taxon:100226"																	
116. .1390																	
/gene="SCO7543"																	
116. .1390																	
/note="synonym: SC8G12.19"																	
116. .1390																	
/gene="SCO7543"																	
/note="SCO812.19, possible Crp-family transcriptional																	
regulator, len: 424 aa. Similar to several including:																	
Bacillus subtilis SW:XMR BACSU (EMBL:M27248) xylose																	
repressor (384 aa), fasta scores opt: 339 z-score: 347.1																	
E(): 7.1e-12 23.3% identity in 390 aa overlap and																	
Streptomyces coelicolor TR:050502 (EMBL:AL009199) probable																	
transcriptional repressor protein, SC7B7.05 (403 aa),																	
fasta scores opt: 691 z-score: 697.8 E(): 2.1e-31 33.9%																	
identity in 387 aa overlap. Contains a Poxite hit to																	
PS01125 ROK family signature and Pfam hits to PF00480 ROK																	
ROK family and PF00335 crp. Bacterial regulatory proteins,																	
crp family with the putative helix-turn-helix motif																	
situated between residues 32..53 (+/-4.61 SD)."																	

```

misc_feature
206..298
/feature="SC07543"
/note="Pfam match to entry PF00325 crp, Bacterial regulatory proteins, crp family, score 15.80, E-value 0.0014"

misc_feature
680..958
/feature="SC07543"
/note="Pfam match to entry PF00480 ROK, ROK family, score 96.80, E-value 1.1e-27"
794..877
/feature="SC07543"
/note="PS01125 ROK family signature"
1485..2288
/feature="SC07544"
/note="synonym: SC0812.20"
1485..2288
/feature="SC07544"
/feature="SC07544"
/note="SC0812.20, possible ABC-transport protein, membrane component, len: 267 aa. Similar to many other transporters e.g. Escherichia coli SW:TAUC_ECOLI(EMBL:D85613) tauroine transport permease protein, Tauc (275 aa), fasta scores opt: 547 z-score: 643.1 E(): 2.3e-28 37.5% identity in 248 aa overlap. Contains multiple possible membrane spanning hydrophobic domains and a Pfam match to entry PF00528 BPD transp, Binding-protein-dependent transport systems inner membrane component."
/codon_start=1
/transl_table=11
/product="putative ABC-transport protein, membrane component."
/protein_id="CAB90902.1"
/db_xref="GI:779280"
/translation="MAADTSAGLAESAPEPERSGRRGLPLVLANVSVALGIALMMWALA SAGFLPPEPVNBARSLIGDGLDADLALSLRVILIGELGAVAVPGFLGWTGVT LRGLIEWIQPFRTIPPLAIIPLAVAMGIDRPRKI FVPLAFLAISTPGCVNNV LDRITNARVLGADATIFARVAVVPASTPTPLGMRVGLSASAMATIVAAELIIAAGCG LGYRNQNLAYDPTLPTIFVGLISIGILGLMDRVLILAERLTKMOERR"
1953..2168
/feature="SC07544"
/note="Pfam match to entry PF00528 BPD transp, Binding-protein-dependent transport systems inner membrane component, score 44.90, E-value 1.8e-09"
2285..3097
/feature="SC07545"
/note="synonym: SC0812.21"
2285..3097
/feature="SC07545"
/feature="SC07545"
/note="SC07545"
/feature="SC0812.21, possible ABC-transport protein, ATP-binding component, len: 270 aa. Similar to multiple ABC-transport proteins e.g. Escherichia coli SW:TAUB_ECOLI(EMBL:D85613) tauroine transport ATP-binding protein, Taub (255 aa), fasta scores opt: 655 z-score: 772.8 E(): 0.434% identity in 235 aa overlap. Contains a Prosite hit to PS00017 ATP/GTP-binding site motif A (P-loop) and a Pfam match to entry PF00005 ABC_tran, ABC transporter."
/codon_start=1
/transl_table=11
/product="putative ABC-transport protein, ATP-binding component."
/protein_id="CAB90903.1"
/db_xref="GI:779281"
/translation="WTSSTPAKISVOHVTKPTSGREFTLLDVSLDIANHPEFTVVG BSGCKSLTNIIILGAPRTSGALVNGAPSGSGPRGVI FOYALFPLTYRNVSE FGLRTGPKERRARABHFIELVGLERFADALKTISGGRKRCATARYANDPSIL EMLDEFGALDLTVKLQEOILKTWSQDRDRIETITDVEAFLARVAVMAAPGR VYDVVDVPAIEREAEARLSPEEPAALRNRRWRSVYHOEGRTAALSGSHSP"
2393..2929
/feature="SC07545"
/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 224.10, E-value 2e-63"

misc_feature

```

```
misc_feature      2414..2437
                  /gene="SC07545"
                  /note="P500017 ATP/GTP-binding site motif A (P-loop)"
gene              3094..4116
                  /gene="SC07546"
                  /note="synonym: SC8G12.22"
                  /gene="SC07546"
                  /note="SC8G12.22, possible solute-binding lipoprotein,
                  len: 340 aa. Similar to several e.g. Bescherichia coli
                  SW:TAUA_ECOLI(EMBL:D85613) taurine-binding periplasmic
                  protein precursor, Taura (320 aa), fasta scores opt: 235
                  z-score: 267.4 E(): 1.9e-07 23.3% identity in 287 aa
                  overlap. Contains a possible N-terminal signal sequence
                  and an appropriately positioned prosite hit to P500013
                  Prokaryotic membrane lipoprotein lipid attachment site."
                  /codon_start=1
                  /transl_table=11
                  /product="putative lipoprotein."
                  /protein_id="CAB90904.1"
                  /db_xref="GI:779282"
                  /translation="MTRKGPMTAMSPVLTFRSLTTLTLAGVAAVLLAACGSSSGSSD
                  KRFPGYINDFNASLIAlAEAGLWKHGLSADAKVFTNGPIQIQLGTDNDPFGIG
                  PAMMLPASGAQKVAIAINTGNSDRVLAOPGITSMOQLKTTAVAEVSGMILSLA
                  LEKAGMTKEDLVKVPMDPSTIVASFSOOVDGAGFVPAATIKTQVPLVLANRAD
                  FEQDVSFPTAFVAGKRVAVDOPEKTKTAVLREANTPSSHTDAIKITAKLKIPV
                  EGVKDAANNKVLVAELDELTRDGTIVKMLSGMDFVNAKGLKKPDPKTYITGDL
                  FTGAGK"
misc_feature      3163..3195
                  /gene="SC07546"
                  /note="P800013 Prokaryotic membrane lipoprotein lipid
                  attachment site"
                  /gene="SC07547"
                  /note="synonym: SC8G12.23"
                  /gene="SC07547"
                  /note="SC8G12.23, possible sulfatase (fragment), len: >136
                  aa, low similarity to SW:YIDU_ECOLI(EMBL:L10328)
                  Bescherichia coli hypothetical 57.3 kDa protein in
                  Emdr-Glyc intergenic region putative sulfatase YdJ1 (EC
                  3.1.6.-), 497 aa; fasta scores: opt: 147 z-score: 185.4
                  E(): 0.0085; 29.6% identity in 115 aa overlap
                  SC8G12.23, possible sulphatase, partial CDS, len: >330 aa.
                  Similar to many Eukaryotic sulphatases e.g. Homo sapiens
                  precursor (BC 3.1.6.13) (550 aa), fasta scores opt: 326
                  z-score: 382.2 E(): 7.9e-14 28.4% identity in 423 aa
                  overlap. Also similar to Bescherichia coli
                  TP:YIDU_ECOLI(EMBL:L10328) hypothetical 57.3 kd protein
                  (497 aa), fasta scores opt: 221 z-score: 0.0 E(): 0.0
                  31.2% identity in 397 aa overlap. Contains a prosite hit
                  to P500523 Sulfatases signature 1 and a Pfam match to
                  entry PF00884 Sulfatase, Sulfatase."
                  /codon_start=1
                  /transl_table=11
Query Match      14.2%; Score 600; DB 1; Length 302007;
Best Local Similarity 52.8%; Pred. No. 3.9e-85;
Matches 1622; Conservative 0; Mismatches 1350; Indels 102; Gaps 11;
```

```
QY      182 TCGTCCCGCCCGACGAGCGGCGACGCTCCGTTTCTCTGACAGACATCCAAATCCT 241
DB      147644 TCGGCGCGACCCCGACACCGACGAACTTCTCCCTGACGACATCCAGCGCGCT 147703
QY      242 ACTGCTGCGCGCGACAGAGAGGTTTAC---GCTCCGAGCGGAGTCCAGCGCTTATCCG 298
DB      147704 ACCCTAGCGCGCGACCGACCGCTACCGCTACGCGGAGGCGTGCAGCGCTACGTGG 147763
QY      239 AATACGACTGATACGATCTTCAACCTGCGGAGCTGAGCCGCGCTTTGGAAAGTCGTG 358
DB      147764 AGCTGCAATATCCCACTTCACCTGACCGAGTACCTCGTGTGGCGGCAATCGTTC 147823
QY      359 CGCGGACGACATGCTTGGCGGCGCCACAGCTGCGCCGACATGATGAGGATGATCGAGCTTA 418
DB      147824 AGCGGACGACATGCTTGGCGGCGCCGATTCACACGAGAGGCGACAGCGGCTCTGCGCG 147883
QY      419 AAGTCGACGC---CGACATCGAGATCATGATCTGCGGCGGCTCGACCGGAGACACAGCG 475
DB      147884 AGTGCCTCGCGCTGTCCTCGTGGGACCGACGACCTGACCGCGCTGCGGCGCACGCGCG 147943
QY      476 AAGCAGGCTGCTGTCGTTCGCAAGATGCGATGTCGACCCGATCTATGACCCGAGCGCC 535
DB      147944 AGGCAACGATGCGGACACCCGCGCGCGGCTGTCGCGCGAGCGCCACCGACAGT 148003
QY      536 CTTCGCTCTATCAGCTGCTGCGCGCTTGGAGTGAAGAGCGGCAACCCGCTCTGCTCTCA 595
DB      148004 GCGCCCTCT--TCACTGATACGTCGACCCGACCGACCGCGCGGCGCTGATCGACCTCT 148060
QY      596 GTATGATCTCATTTAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 655
DB      148061 CTTTGAACATGCTGCTGCTGACACGACGACGCTGCTGCTGCTGCTGCTGCTGCTG 148120
QY      656 GCTTTTGAAGATATCCGAGACCTCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 715
DB      148121 GGAGCTACGCGCGACCG---CCTTACCGACCGCGCGCGGATCATCTTTCGCGCTACG 148177
QY      716 TACTGCGCTGAGTCTGCGAAGATGTCAGGCGCATCAACGATGATGATGATGATGATGAT 775
DB      148178 TCTTGGCCCTCGCGCGCGCTTGTCCGACACGACGCGACCGCGCGCTTACTGGA 148237
QY      776 AGCGGCGATGCGCGAGCTCCACCTCC----- 803
DB      148238 CGGAACGCGTGAACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148297
QY      804 -----GCGGACGCTTCCATGAAAGCGCATCATCTACCTGTAAGAGATCCG----- 851
DB      148298 CCGCGGTGACGACCCGCGGAGGCGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 148357
QY      852 -CTTCCGCGACACGAGCAATGCTGCGTGGAGCTCTCTGCGGAGTGAAGCGCGCTG 910
DB      148358 CGTTCCGCGCGTGGAGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148417
QY      911 TCGGGGAGCGCGGCTGACCCCGGAGCGGCGTCACTTGGCTGATTTTCCGAGTGAATG 970
DB      148418 CCGCGCGCGCGCGGCTGACCCCGGAGCGGCGCTCTCAACCTTACCGCGGAGAGGCTG 148477
QY      971 GCGCGTGAAGCGCGAGCCCGCGCTTACGCTCAATACGCTTTCACCGCGCTCCCG 1030
DB      148478 GCGCGTGTGCGGCTGACCGCGCTTACCTTAAAGTGCACCGCTGACCGCGCGCGCG 148537
QY      1031 TCCATTCGCGCGTGAAGCATATCACCGGAGACTTCACTGATGATGCTCTCTGACATG 1090
DB      148538 TGCACGAGGACATGACCGCGCTCTGCGGAGATTTACCTCGTGAACCTCTCGCGCG 148597
QY      1091 ACACCACTCGGACAAAGAGTTTGAACAGCGCGCTTAAAGCTTAAAGAGACGCTGTGG 1150
DB      148598 ACCTGACACACCGCGACCTTTCGCGGAAAGAGCGCGCGCGCTGCGGACAGCTCTCG 148657
QY      1151 AAGCATGATATACCTGAGACGTAAGCGATATGAGATGATGAGGAGAGAGCGCGCGGATCC 1210
DB      148658 ATGACCTGCGCCACCGCGCTTTCACCGGCTGGAAGTGTG-----TGGCGAATCTCTCC 148711
QY      1211 TGGGATCAACGAGGCGGATTTTCCCGTGTGCTCAAGAGCGCGTTTAAACAGCAAG 1270
```

Db	148712	GCAGGGCGCGCGCTCCGGTCCCTCATCCGGTGGTCTTCAACAGCGCCCTG-----G	148762
Qy	1271	TCGTTGGTGTCACTTCCTGTTGCAGAGGCTCGAACTTCGGGTATACACAGACGCAATTC	1330
Db	148763	GGCGGGGGCCACTCCCGAGGGCGTAACCCCGGAGGTGAGTACGCGTGAACCCGAGCAC	148822
Qy	1331	CTCAGCTCTGTGTGATCATCAAGCTCTACAGACGATGGGGACCTGTCCTCGGTGGG	1390
Db	148823	CGCAGGTGGCTTCACCTGACCAGTATATCAACGCGGCAACCTTCAGCTTCTCTGGG	148882
Qy	1391	ACATGTCGACGGAAGTGTTCGCCCGCACTTCTGAGCAATGCTGAAGCTAAGTCG	1450
Db	148883	ACATCGGGAAGGGCGGCTGGCCACAGGAGCGGACCGCATGTTCTGAAGGCTAACGG	148942
Qy	1451	TTTTTCTCCGGGGCTCACT-----GAGAAACATGGGGTGAAACAGG	1492
Db	148943	CACCTGTCGTTCTGCTGTGCGGAGGCGAGGCGAGGACAGAAAGGCTGGAGCGACCGG	149002
Qy	1493	TGCGCTGTTCGTTCTTCGCTCGCCAGCTGAAGACGGCGGCGAGCGCAAAACGCAACAAG	1552
Db	149003	TCCGATCCCGCTGCGCGGAGACAGGGGCAAGCGGCAAGCGGCGCGGTCAAGCGAACGGAG	149062
Qy	1553	CGCTGCTGAGCGAGCATACGCTGACAGGCTGTTGCGGGCGGGGTGAGCACTGCCA	1612
Db	149063	GCCCCCTCCCGGACCCCTGCTTCACGAAACCGTACTCCCGGGGACGGAACCAACCGG	149122
Qy	1613	TGACGCTCGCGGTGGTGTGCGGCGCAACGCTACGTACGAAAGCTTTGCGCGGTT	1672
Db	149123	ACGCGATGCGGCTCCGCAACCCCGAACTGGCTCTCTTACCGGACAGCTGTACCCGG	149182
Qy	1673	CGCGGCACTTGGCGCGCGGCTGCGGAGACAGGGGACCGCCGAAACAATGGTCGGG	1732
Db	149183	CCACAGGACTCGCCAGCACTGACGGCTCCGGATCTGGCGCCCGGGGAAACGGTGACCA	149242
Qy	1733	TGGTATGAGAAAGCTTGGGAGCAGTGTGTGGCGTTCTCGGGTGTCTGAGTCAAGGG	1792
Db	149243	CTGAGATGACAAAGGCTGGGAAACAGTGTGTGCGCTCTTTCGACACTTCATAGCGGGG	149302
Qy	1793	CGGCTTACGTGCGATTCGATGCGCACTACCGGCGAGGTATCCACTACTCTTGATC	1852
Db	149303	GCGCTTACTGTGCGGTTCGACAGCGGCCACCGCGGCGCGCGGAGACAGTATATGGGG	149362
Qy	1853	ATGTGAGGTAAAGCTCGTGTCTGACGAGCCATGTGCTGATGGCAATCTGTATGCGCG	1912
Db	149363	ACGCGGGGCTCGGCAACCGTCTTCAACCACTCGGTGCGCGGAATCGAAGACTTCCCT	149422
Qy	1913	CGGGGATCCAGCGGCGCTCGTGAAGAGGACGGGCTCGAAGGCGACCGGACACAGCTTC	1972
Db	149423	CGACGCTGTCCCGGTGCCGT-----GAACTCGTGGGAGGCAACCGGACCTTCC	149476
Qy	1973	CGATGATGCCCAATTGAGACACTTGGGATCTCGCGTATGTATCAACTCGGGATCCA	2032
Db	149477	CGCCCGCGCGCGGAGCCTTGAGACCTCGCTTACGTATCTTACACTCGGGCTCCA	149536
Qy	2033	CAGGTTGCCCAAGGGGGTGAATGATATCTGGGGTGCCTGTACAACATCTTGACA	2092
Db	149537	CGGCAACCCCAAGGCGGTGATATGACCAACCGGGCGCCCTTCACACCGTGAAGACA	149596
Qy	2093	TCAACGACGCTTTCGAAATAGGGGCGGAGACAGGGTGTCTGGGCTCTCTCGCTAGCT	2152
Db	149597	TCAACCGCGCTTTCGCGGTTCGACGAGCGGACCGCGCTCTCGGCAATCGCCGACTCGCT	149656
Qy	2153	TCGATCTCTCGGTCTATGATGTGTTTGGGATCTCTGGCGGCGGGCGGATCGATGTGGTGC	2212
Db	149657	TCGACTCTGCTCTTACGACTGTGTTGGAACCGCTCCCGTGGGCGCACCTCTGTCTGC	149716
Qy	2213	CGGACGCGTCCAAAGCTGCGCGATTCGGGCGATTGGCAGAGTATGATCGAACGAGAAAG	2272
Db	149717	CCCGGTCCGACTGTGGCGGCGACCTTCCCACTGGGCGCAATCTGTTCGGAATTGGGG	149776
Qy	2273	TCACGCTGTGAACCTCGGTGCGCGCGCTGATGCGATGCTGTGAGCAATTTTGAAGGTC	2332

Db	149777	TCACGAGTGTGGAACTCCGATCCGGGCGAGCTCACAATGCTGTGGCACTGGCTCCGCG	14983
Qy	2333	GCCCGAATTCCGCTCGCTAGTGTCTCTGGGCTTTCCGCTGCTGAGCGGCACTGGATCCCGG	2392
Db	149837	AGCCCCCCACCGACACAGCGGCTCTCCCTGGCGCTGAGCACTGATATCGGGCGAGCTGATACCGG	149896
Qy	2393	TGGGCGCTGCTTGGCAGGCTCCAGGCGCATTCAGGCGCGGCGCTGTGCGTGAATCAGCGTGGCG	2452
Db	149897	TGCGACTCTCCGACCAAGGCCCTGTAACTCTGCCCCGAGATGGAGATCGTCTCCCTGGCGG	149956
Qy	2453	GAGCCACCGAAGCGTCGATCTGTGTTCATTCGGGTACCCCGTAGAGAACTGTGACTATCGT	2512
Db	149957	GAGCCACCGAGGGCTCTCAATCTGTGTTCATTCGCCACCCGATCGGCGAGGTCTGACACCGCCC	150014
Qy	2513	GAGCGAGGATCCCTTACCGGCGCGTCCGCTGCGGCAACGACAGCGTTTCCAGCGTCTGATGAGG	2572
Db	150017	GCCCCGATCCCGTACCGGCAAGCCCTGTGACCAACAGACGTTCCGCTGATCTGACCGGC	150078
Qy	2573	CGCTTGAACCGGCGCCGCTCTGTGGTTTCGGGGCAACTTACATTTGGCGGGGTTCGGGCTGG	2632
Db	150077	ACCTGCGGCGCGGGCGCCCGAGTGTGTTCGGGGGAGCTGTACATCGGCGGGTGGCGGTGTGCG	150136
Qy	2633	CACGTGGGCTTACTGGGCGGATGAAAGAAAGACCGCGAGAGCTTCTCTGTGCAACCCGAGA	2692
Db	150137	CCCTGTGGTACTCTCGGCGACCGGCGAGCGCACGACACGTTCTTCACCGACTTGGCCA	150196
Qy	2693	CCGGGAGACGCGCTCTTACAAAGACCGGCGATCTGTGGCGCTACTGCTGCCGATGGAACATCG	2752
Db	150197	CCGAGAGACGCGCTCTACCGGCAAGGGCGACTGCGCGCTACTCTGCCCCGAGCGGAGCATCG	150255
Qy	2753	AGTTTATGAGGCGGTGAGGACCAACAAATCAGCTTTCGCGATACCGCGTTAGCTGGGG	2812
Db	150257	AGTTCTCTGGGCGGAGGAGACGCGCCGATCAATTCGCGGATACCGGGGTCTGAACTGGCGG	150314
Qy	2813	AAATGAGAGAAACGCTCAAGTGCATTCGATCGGAAGTACGCGACGCGGTGATTTGTCCCGTCG	2872
Db	150317	AGGTGAGAGCGCGCTTACAGATCCACCCTCGCGGTGCGCGGCGCGGTCTGTGTGAGCG	150376
Qy	2873	GGAACGACGCGCGCGAACAAGCTCTTCTTAGCTATGTGTCCCGAGGGGACACGAGAC	2932
Db	150377	ACTGCGGCGCGGCGGCGGCGGCGGCGCTGTGGCGCGTGTGTGAGAGCGACCAAGACGCGCG	150438
Qy	2933	GCGCTGCGGAGCAGACGACCGGACCTTCAGACCCGAGCGGATTCGACCGGAGAGCACACGCGG	2992
Db	150437	GCGCCGCGCGGTGACGAGCGCGCGACAGGCGCGCACGCGACCGCCGAAAGCGGTACGCGAAG	150496
Qy	2993	CCGAGCGGACGCGC 3006	
Db	150497	CGGCGCGCGCGCTC 150510	
RESULT 15			
AL646066			
LOCUS	Ralstonia solanacearum	189050 bp	DNA linear BCT 11-APR-2003
DEFINITION	Ralstonia solanacearum GMI1000 chromosome, complete sequence;		
ACCESSION	AL646066 AL646052		
VERSION	AL646066.1 GI:17428714		
KEYWORDS			
SOURCE	Ralstonia solanacearum		
ORGANISM	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;		
REFERENCE	Burkholderiaceae; Ralstonia.		
AUTHORS	1 Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,		
	Arlat, M., Billault, A., Broctier, P., Camus, J.-C., Cactolico, L.,		
	Chandler, M., Choisme, N., Claudel-Renard, C., Cumac, S., Demange, N.,		
	Gaapin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,		
	Siguier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,		
	Weissenbach, J. and Boucher, C.A.		
TITLE	Genome sequence of the plant pathogen Ralstonia solanacearum		
JOURNAL	Nature 415 (6871), 497-502 (2002)		
MEDLINE	21681879		

```

PUBMED 11823852
REFERENCE 2 (bases 1 to 189050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, UMG CNRS
118 Route de Nardonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT http://sequence.toulouse.inra.fr/R.solanacearum.html.
location/Qualifiers
FEATURES
source 1..189050
/mol_type="genomic DNA"
/strain="GM11000"
/db_xref="taxon:305"
170..3325
/gene="RSC1698"
/note="synonym: RS02885"
170..3325
/gene="RSC1698"
/function="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE BACTERIOPHAGE PROTEIN"
/protein_id="CAD15400.1"
/db_xref="GI:17428715"
/db_xref="SPTREMBL:O8XY08"
/translation="MRNIIYGKGGKSSSTPTPEADPSLHSVAVRVLVSEGEIA
GLVNGLSIYLGSTPLANNADGTLNFONVADVDRPGSODEALRGPPSVESAEVLEL
TATMPWVAIANTOLSAVRVOLSVALSPADSSCNINGRVAVLIDLAVGALAOV
LATADGKTSKTRTRHRELPRAKTGTIRRTIRPNASGTLADVTRIEIAVIO
ALRYPNLSALIGRIDARQSSVPTRSYHLRGRITRPENYDPAIRTPVSPVOAFT
ASNNPAWFEYDVIHKGGLDRVSADVDWMALYQIGYCELVPPDRGQBERFT
NCYCIQORNDAPVLQDLASVFRGMAYMSNVAVADMPASAYLFHGNVDRERFT
YAGSARBARKTVALVSMNDPGQYRAKVEHVEDGGIARISIOOTEVAFGCTSOOA
ORVGOMILTSRLTEIETVSFKYGLDAVVMPGAIIENVADPARAGTNGRVRASDART
VYVDRPVYVRESDTILINLPDGTAAORTISRVDGSLTYSADMSQPVQAEAVMSYESA
DKITQLFRVSVTESDGLTFDISALEYNOSKPSALDNGRTRIRARPVSVLPSPVAFPT
DYTLATYSVVDGIAITTMVMVAMQQAASAVALATVEMRDRNGEMVTAAGRVSQSLERN
IYAGTYVAVRVAIINALDVASLPYSAEFTLTKTSPPVVGTILATATVFGIRLMAF
PTGPDVERTELMSKAPRPSATKLGDCAPANTHTMLAAGATFFYMAALVDSKG
NVCAYVPSGNGVIGMSDPAADLSIYMGKDFATLGLDILSAVDSIDPPMAGSEDF
AGDGYVAGIVSTQSVLEARNRAIKQOTTMATYAQSSAAVQAQAAVABEGGILAA
MTTIQTQIANGRITVLAGIFGVENDKVTESQVLIADRFAVLHPNGNNVTPTPVQ
GGQVPMDSAFIADGTTISAKIDGSIQSSNFVAGQYGMRLSKGLFENNNGAGGRVD
TSLATQIYDANGTLRLGVW"
3662..4318
/gene="RSC1699"
/note="synonym: RS02888"
3662..4318
/gene="RSC1699"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
gene
4387..4878
/gene="RSC1700"
/note="synonym: RS02889"
4387..4878
/gene="RSC1700"
/EC_number="3.2.1.-"
/function="elements of external origin; phage-related
functions and prophages"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE BACTERIOPHAGE ENDOLYSIN PROTEIN"
/protein_id="CAD15402.1"
/db_xref="GI:17428717"
/db_xref="GI:17428717"
/db_xref="GOA:O8XY08"
/db_xref="SPTREMBL:O8XY08"
/translation="MPYTDPSVLGGRNLAFLDMVGFSEGTNGRQPTRDGVDLVG
GLFVSYADHRIIVLELPLGIKSTAGRYOLLARYDPPYRROLVDFPGAADALIA
VQIIRERCALDIDQGRLLAAIACKNIMASLPBAGYQHCKREBALRAHLLCCGGE
VKG"
4878..5186
/gene="RSC1701"
/note="synonym: RS02890"
4878..5186
/gene="RSC1701"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE LIPOPROTEIN TRANSMEMBRANE"
/protein_id="CAD15403.1"
/db_xref="GI:17428718"
/db_xref="SPTREMBL:O8XY07"
/translation="WTSEBEMVAAKVGSALGSVIALRFLPTWQRMLSFVSSLGI
GLNGGAVERFLLVPSGYTHMLAVASAAVGLAIWNNAQOIPILTLDRRFLAKE"
5189..5464
/gene="RSC1702"
/note="synonym: RS02891"
5189..5464
/gene="RSC1702"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD15404.1"
/db_xref="GI:17428719"
/db_xref="SPTREMBL:O8XY06"
/translation="MLTISINOVANAIIFVGSLWAVLTHQVPTRTGALVLAVNFAA
LGNIAVAGTCHSWPEVALNVAVALCVMAFWRLREVRFRFSALGKRG"

```

gene 5464. .6009  
/gene="RSC1703"  
/note="synonym: RS02892"  
CDS 5464. .6009  
/gene="RSC1703"  
/EC\_number="3.4.-.-"  
/function="elements of external origin; phage-related  
functions and prophages"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Framed"  
/codon\_start=1  
/evidence-not\_experimental  
/transl\_table=1  
/product="PROBABLE BACTERIOPHAGE-RELATED SIGNAL PEPTIDE  
PROTEIN"  
/protein\_id="CAD15405.1"  
/db\_xref="GI:17428720"  
/db\_xref="GOA:O8XY05"  
/db\_xref="SPTREMBL:O8XY05"  
/translation="MPGKGTATIAATIMILGALVGAAGWTMCAHYRDKDIALFLORD  
QAEKDROQARADENQRLREKREKRLERQLELDARHRELNKTERARAAALAG  
TRQLSIRACTAPADGVPSITSGAGLGDGAATAILGEDAD1QALVADAD1AAL  
RFQGRERAIQTRDTSVDAR"  
complement(6117. .7304)  
/note="predicted by Homology"  
/insertion\_seq="ISR018 or RS05932"  
complement(6273. .7238)  
/gene="tISR018"  
/function="elements of external origin; transposon-related"  
repeat\_region  
gene  
CDS  
Query Match 14.0%; Score 593.6; DB 1; Length 189050;  
Best Local Similarity 52.6%; Pred. No. 4.3e-84;  
Matches 1510; Conservative 0; Mismatches 1309; Indels 51; Gaps 8;  
206 ACGCTCGTTCCTCTCAGACATCCAGAACTCTAGTGGCGCGGACAGAGAGCT 265  
Db 159271 ACGACACTTCGCTGACCGACATCCAGACGCTTACTGATCGCGCGGACAGGCTC 159330  
Qy 266 TTACGATCCCAAGCGG--ATCCAGGCTTATCCGGAATAGACTGTAGCTCGACG 322  
Db 159331 TGGGGCTCGGGGCTGTCATGCAACGCTTACTTGAATGGGTCTCGCAGACCTGAC 159390  
Qy 323 TGGCAGGCTGAGCGCGGCTTTTGGAAAGTGTGCGCGGACGACATGCTTCGGGCCC 382  
Db 159391 TGCACGCTTCGACGAGGCTGGAACCGGATCGTGGCGGACGACATGATGCGGCGC 159450  
Qy 383 ACAGGCTGCCGACATGATGAGAGATCGAGCCCTAAAGT---CGAGCCGACATCGGA 439  
Db 159451 TGCTGACCCCGAATGAGGCGAGCGCATCTCGGACCGTCCGACATACGCCATCGGA 159510  
Qy 440 TCATGATCTCGCGGCTCGACCGGAGACACGAGGAAGAGAGCTCGTGTGTTGCGAG 499  
Db 159511 CCGAAGACTTCGCGCGGACCGGACCGGATCGGCTGCGGCGGACGAGCTCGCGAAGACCG 159570  
Qy 500 ATGCGATTCGACCGCATCTATGACACCGAGCGGCTCGCTTATCAGCTCGTCCGCG 559  
Db 159571 CCGGCAATGCGGGAGGTGTCTGACGCGGAGCGGTGCGCTGTTCGACCTCGGCGGCA 159630  
Qy 560 TTGCGCTGAGAGAGCGGAAACCGCTCTGCTGCTCAAGTATGATCTCAATTAACGTTGAC 619  
Db 159631 CGCGAGACCGGCGCGGCTGACCGGCTGACCTTGATCTGATGATGTTTCGACG 159690  
Qy 620 TAGCAGCTGCTCATCTTTCAAGACTGAGCTTCAAGAGATCCGAGACT 679  
Db 159691 TGCGAGGTTCCATCTCTGCTGCGGCGGAGCTCGAGGAGACTACCGGACCGCGCGG 159750  
Qy 680 CTCTCCCTGCTGAGCTCTGTAACCGCATATGTACTCGCGCTGAGTCTCGCAAGA 739

Db 159751 AGCTGCCCGCCATCGGCTCAAGTTTCGAGACTACTGCTGGCGCGGAGACCGCGG 159810  
Qy 740 AGTCTGAGGCGCATCAACGATCGATGATTAAGTGAAGCGGCGATCGGAGCTCCAC 799  
Db 159811 GCGAAGCAGCTTATCTGGCCGACAGAGCTGTGTGGAGCCCGCTGCGAAGATCGGC 159870  
Qy 800 CTTCGCGGACGCTTCGATGAGAGCGGATTCATCTTACCTGGAAGATCGCTTCGCGC 859  
Db 159871 CGCCCCGCGGCTCCGCTCGCGGCGGTGGCGCTCGAGCTGGCGCGGCGCATTCGCGC 159930  
Qy 860 ACAGGAGCAATGCTCGCGCTCGGACTCTCGGGGTGATTAACCGGCTTCGGGAGC 919  
Db 159931 GCTTCAGCATCGCTGACCCACAGGCTGGCGAGCGGCTGTGCTTTCGCCCGAGG 159990  
Qy 920 GCGGCTGACCCCGACCGGCGCTCATCTGCGCTGATTTCCGAGTATCGGCGCTGGA 979  
Db 159991 CCGGCTGACCGGCTCGTCCGTGCTGCTCAACCGGCTTCGGAAGTGTCCGCGGCTGG 160050  
Qy 980 GCGGAGCGCCCGGTTTACGCTCAACATTAACGCTTTCAACCGGCTCCCGCTCATCGC 1039  
Db 160051 CCGGCGACGGCGGCTTACGCTCAACCTCAACGCTTCAACCGCGCGCGCTGATCCGG 160110  
Qy 1040 GCGTGAACGATATACCGGGGACTTCACGCTGATGCTCTCTGAGACATGACACATC 1099  
Db 160111 ATGTGAGCGGCTGTGCGGCTTCACTCGGTGCTGCTCAACGCTCACTGCAACG 160170  
Qy 1100 GCGACAGAGCTTGGAACAGCGGCTTAAGGCTATTCAAGAGAGCTGTGGAGAGATG 1159  
Db 160171 CGCGCTGCTTTCGCGCGGCTGACGCTGCGCGGACCGGACCGGACCGGACCGG 160230  
Qy 1160 ATCACTGCAAGTAAGCGGATATGAGGTTCAGCGAGAGCGGCGG-----1205  
Db 160231 CGCACAGCGGCTTGGCGGCGGCTGGAAGTCTGCGGCAACTGCGCGGACCGGCGGAG 160290  
Qy 1206 -GCTCTGGGATTCACAGAGCGGCTATTTCCCGTGTGCTCAACGAGCGGCTTAAC 1264  
Db 160291 ACGGCGGCGGATGAGCGGCGGCGGCTGATGCGGCTGTGCTTCAACGAGCTTCG 160350  
Qy 1265 AGCAAGTGG--TTGATGTCACCTGTTGCGAGAGCTCGGAATCCGGTGTACACGCA 1321  
Db 160351 TGATCTGACCTCTGCTGCTGCTGCGGCGGACCTGCTGCGGAGCGGAAATTCCTTACA 160410  
Qy 1322 CGCAGACTCTCACTGCTGCTGATCATGCTTACGAGACAGATGGGAGCTCTGTC 1381  
Db 160411 CTGTCAGCGCCAGGCTGCTGCTGACACAGAGATGATGCTCGAAGGCGACGCTGGAAT 160470  
Qy 1382 TCGCGTGAACATGTGTGAGCGAAGTTCCTCGCGGACCTTCTGAGACATGCTCGAAG 1441  
Db 160471 TCACTGATGATGATGATGAGCACTGTTCCCGGAGGCGATGTGATGCTGATGAG 160530  
Qy 1442 CGTACGTGTTTTTCTCGCGGCTCACTGAGG-----ACATGGGGTGAACAGGTG 1495  
Db 160531 CTATTTGCGGCGGCTGACCATCTGAGCCAGAGATGCCGCGATTTGCGAGCGCATGTC 160590  
Qy 1496 GCTGTTGCTTCGCTGCGCAGCTAAGAGCGCGGCGAGCGCAACGCAACGCGC 1555  
Db 160591 CTGCTGAATTCGCGGAGCGAGCGCGGATGCGGAGGCGCTGAACCGGACCGCGCTAC 160650  
Qy 1556 TGCTGAGCAGATACGCTGACGCGCTGTCGCGCGCGGCTGAGACGCTGCCATGC 1615  
Db 160651 CGCTGCGGAGTCCGCTGACCGGAGCTTCTTCGCGGAGGCGGATGCGGACGCGCACG 160710  
Qy 1616 AGCTCGCGGTGTTCGCGCGGCGCAAGCGCTCAAGTACGAAAGACTTTCGCGCGTTCG 1675  
Db 160711 CGTGGCGTTCGCGACCGGCTATGCGAGAGACTTATGCGGACTGAGCGCGCGCGC 160770  
Qy 1676 GCGACTTTGGGCGCGGCTGCGAGGAGGCGGCGACCGGCAACATTTGCTCGCGGTG 1735  
Db 160771 TGGCGCAGTTCGCGACGTTTGGGAGGCGGCGGCTGATGCGCGGAGTCTGTTGCTGCG 160830  
Qy 1736 TGATGAGAAAGCTGAGAGAGCTTGTGCGGTTCTGCGCGTGTGAGTCAAGCGCGG 1795  
Db 160831 CAGTCCCAAGGAGATTCGCGGCTGCGCGGCGCATGCGGCTGCGCGGCGGCGG 160890



```

QY 1796 CTTAGTGGCCGATTCAGTGGCCGACCTTACCGGCGGAGGTATCCACTACTCTCGATCATG 1855
    |||||
DB 160891 TCGTGGTGGCGTTGGGCGACGACATGCCCCCGCGGATGAGCCATTTCTCGATGGCG 160950
    |||||
QY 1856 GTGAGGTAAAGCTCGTGTGACGACGACATGGCTGGATGGCAAACTGTATGACCGCGCG 1915
    |||||
DB 160951 CAGGTGTCAAAAGCCGCTTTATCGCGGACGAGCGGACTTACGGGCGCGGATGCGCAAGC 161010
    |||||
QY 1916 GGATCCAGCGGCTGCTC-----GTGAGCGAGGCGCGGCTTCGAGGCGACGGCGACAGC 1969
    |||||
DB 161011 CGGTGGCGGTGGGTGTGTCGCGGAGTCCGAGCGGCGCGTGTACGCAAGCGCGCGCTGCG 161070
    |||||
QY 1970 CTCCTGATGATGCCCCATTGAGACCTTCG-----GATCTCGCGTATGTCATCT 2017
    |||||
DB 161071 AGCGGAGATTCAGACGGGACGTGCGCTCCCGCCTCGATGCGCTGCGCTTACATCATCT 161130
    |||||
QY 2018 ACACCTCGGGGATTCACAGGGTTGCCCCAAGGGGTATGATCATCATCGGGGGTGGCGTCA 2077
    |||||
DB 161131 ATACTTCGGGCTTGACCGGCGACGCCCCAAGGGGCGTGGCGATTCAGCGACGGCGCGCGGCA 161190
    |||||
QY 2078 ACACCATCTGAGCATCAACAGACGCTTGAAATAGAGCCCGAGACAGAGTGTGGCGC 2137
    |||||
DB 161191 ACACGCTGGAAGCTGTGACGCGGCTGTGGGCTGTGACCGCGACGCGGTGTTCGCA 161250
    |||||
QY 2138 TCTCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTCGGGATCTTGGCGCGGCGCG 2197
    |||||
DB 161251 TCTCGGCGGTGGGCTTTCGACTGTGCGCATGTCTTTCGCGCATGTGTTCGCGGCGCG 161310
    |||||
QY 2198 GTACGATCGTGGTGGCGGACGCGCTCCAGACTGCGCGCATTCGGGCGATTGGGCGAGTTGA 2257
    |||||
DB 161311 GCGCGCTGATCTCTCCCGGACGAGGCGGGGATGCGCATGTGCGACTGGCTGGACTGG 161370
    |||||
QY 2258 TCGAAGCAGAGAGAGGTGACGCTGTGGAATCGGTGCGGCGCTGATGCGGATGCTGTCG 2317
    |||||
DB 161371 CGCACCGGTACGCGGCTGACGATCTGGAATCGGTGCGGCGCTGCTGGCGATGCTGTGCG 161430
    |||||
QY 2318 AGCATTTTAAAGGTGCGCCCCGATCGCTGCTAGGTCTGTGCGGCTTTTGGCTGTGAGCG 2377
    |||||
DB 161431 AGCGCGGGGATTCACAGCGGTTCGGGCGCTCCGCGCATGTGTGGCTGGTCAATCTCAGCG 161490
    |||||
QY 2378 GCGACTGATTCGCGGTGGGCTGCTGCGGAGCTCCAGGCGCATCAGGCCCGGCGTGTGCG 2437
    |||||
DB 161491 GCGACTGATTCGCGGTGGGCTGCTGCGGAGCTCCAGGCGCATCAGGCCCGGCGTGTGCG 161550
    |||||
QY 2438 TGAATCAGCTGAGCGGCGGCGCACCGAAGCGTCAATCTGATCAATCGGTAACCCGTAAGGA 2497
    |||||
DB 161551 TGGCGGCGCTTCGCGGCGCGCACCGAAGCGCGATCTGTGAACTGTTCGAGGTGCGGC 161610
    |||||
QY 2498 AGGTGACCTATCTGTGGGCGAGATCCCTTACGCGCGTCCGCTGCGCAACGACGTTCC 2557
    |||||
DB 161611 GGGTGTGATCCGCTGCGCTGCTCATACGATACGGGTTTCGCTGCGCAACGATCTTACC 161670
    |||||
QY 2558 AGGTGCTGATGAGGCGCTGCAACCGCGCCGCTGAGGTCCGCGGCGCAACTCTACATTG 2617
    |||||
DB 161671 GGGTGTGATGAGGCGCTGCAACCGCGCGATTTGCCCGAGTGGTCAATGCAATGCG 161730
    |||||
QY 2618 GCGGGGTGCGGCTGCGCTGAGCTGAGCGCTACGCGGATGAGAGAAAGCGCGCAAGGCTTCC 2677
    |||||
DB 161731 GCGGCGCGCGGCTGCGCTGTGCTACTGCGCGCATCCGACGCGCACCGCGGAGCTTCA 161790
    |||||
QY 2678 TCGTGAACCCCGAGACCGGGGAGCGGCTTACAAAGACCGGCGATCTGGGCGGCTACTGCG 2737
    |||||
DB 161791 TCAACCGACCCCGCAGCGAGCGCTGTACCGCACCGGCGACCGCGCGCTACTGCG 161850
    |||||
QY 2738 CCGATGAACAATCGATTCATGAGGCGGTGAGGACCAACCAATCAAGCTTTCGCGGATACC 2797
    |||||
DB 161851 CCGACGCGCTGATCGATTCCTTCGCGCGGTGCAAGCCAGTCAAGCTTCGCGGCGCACCC 161910
    |||||
QY 2798 GCGTTGAGCTCGGGGAAATCGAGAAACGCTCAAGTGCATTCGAAAGTACCGGACGCGCG 2857
    |||||
DB 161911 GCATCGAGCTGGGCGAGATCGAAGCGCGCTGACGACGACGAGCGCATTCAGGCGCGCG 161970
    |||||

```

```

QY 2858 TGATTTGCCCCGTGGGAGACGACGCGCGGCAACAGCTCTTCTAGCTATGNGTCCCCG 2917
    |||||
DB 161971 CCGCCAGGTCTGCGCCACAGCGGCGGGGAGACAGCACTGAGCCGCTTCTGTGTCGCC 162030
    |||||
QY 2918 AGGCGACACGAGAGACGCGCTGCGGACGAGACGCGAGCTTCAAGACCGAGCGGATCGAG 2977
    |||||
DB 162031 GACACGCCCCG---GTTCTGCGGACGGTGCGCGCACGATGCCGCTTGGCAGGCGCTGG 162087
    |||||
QY 2978 CGAGACACACGCGCCCGCAAGCGGATGAGCGGATGAGCGGAGAGGCTG 3027
    |||||
DB 162088 CGGCGGCTTGGCGGTTCGCGGCGGCGGCGCTGCGCGCGCGCGCGCAAGGCTG 162137
    |||||

```

Search completed: October 4, 2003, 03:30:57  
 Job time : 15178.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:01 / Search time 2488.11 Seconds  
(without alignments)  
12307.996 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_62369\_63628

Perfect score: 1260  
Sequence: 1 atgcacagagacagcagca.....agccctccaaagctgacag 1260

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_ges\_hum:\*  
18: em\_ges\_inv:\*  
19: em\_ges\_pln:\*  
20: em\_ges\_vrt:\*  
21: em\_ges\_fun:\*  
22: em\_ges\_mam:\*  
23: em\_ges\_mus:\*  
24: em\_ges\_pro:\*  
25: em\_ges\_rtd:\*  
26: em\_ges\_phg:\*  
27: em\_ges\_vrl:\*  
28: gb\_ges1:\*  
29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	9.1	1285	29	B2567880 pac82-164
2	84.6	6.7	975	29	B2675691 PUBDH78TD
3	63	5.0	330	28	AY174775 AY174775
4	55.4	4.4	1127	29	B2548452 pac81-60

5	54.8	4.3	925	29	CNS0091P	AL053013 Drosoph11
6	54.2	4.3	1258	29	B2573154	B2573154 msh2_2968
7	52	4.1	609	12	BI716406	BI716406 1031010A1
8	51.8	4.1	925	29	CNS0091P	AL053013 Drosoph11
9	50.8	4.0	425	10	BG488189	BG488189 RH122_60
10	50.6	4.0	1170	29	B2569411	B2569411 pac82-164
11	50.2	4.0	512	14	CD045174	CD045174 pAHB018X
12	50.2	4.0	512	14	CD045351	CD045351 pAHB019X
13	50.2	4.0	596	14	CD260296	CD260296 pAMA008X1
14	50.2	4.0	610	14	CD043612	CD043612 pAHB009X4
15	50.2	4.0	622	14	CD262676	CD262676 pAMA018X
16	50.2	4.0	715	14	CD041124	CD041124 pAHB001XJ
17	49.8	4.0	1242	29	B2579496	B2579496 msh2_6381
18	49.6	3.9	638	14	CD486072	CD486072 CFEUS_4D0
19	49.6	3.9	691	10	BES59050	BES59050 HV CEB002
20	49.4	3.9	932	29	CNS00720	AL066742 Drosoph11
21	49.2	3.9	402	10	BGS59249	BGS59249 RH122_52
22	49.2	3.9	665	12	BM370205	BM370205 ERT08_50
23	48.8	3.9	788	14	CB668866	CB668866 OSJNEI60
24	48.4	3.8	899	13	BX360973	BX360973 BX360973
25	48	3.8	641	9	AU092915	AU092915 AU092915
26	47.8	3.8	572	14	CA404475	CA404475 E01N0518
27	47.8	3.8	583	10	BE761796	BE761796 947006H01
28	47.8	3.8	799	29	B2977576	B2977576 PUGIS17TD
29	47.8	3.8	838	29	B2977576	B2977576 PUGIS17TB
30	47.2	3.7	389	10	BE229868	BE229868 99AS102 R
31	47.2	3.7	406	9	AU063730	AU063730 AU063730
32	47.2	3.7	413	9	AU093702	AU093702 AU093702
33	47.2	3.7	417	14	D22231	D22231 RICC10540A
34	47.2	3.7	421	9	AU069623	AU069623 AU069623
35	47.2	3.7	442	9	AU182095	AU182095 AU182095
36	47.2	3.7	442	9	AU182164	AU182164 AU182164
37	47.2	3.7	446	9	AU182068	AU182068 AU182068
38	47.2	3.7	449	9	AU161375	AU161375 AU161375
39	47.2	3.7	549	10	BF728783	BF728783 1000067D0
40	47.2	3.7	643	9	AW585283	AW585283 707094C03
41	47.2	3.7	707	9	AU161432	AU161432 AU161432
42	47.2	3.7	719	12	BM600174	BM600174 170006870
43	47.2	3.7	738	14	CB635178	CB635178 OS1EB15E
44	47.2	3.7	741	14	CD428040	CD428040 ETH1_32 D
45	47.2	3.7	764	14	CB668867	CB668867 OSJNEI60

#### ALIGNMENTS

RESULT 1  
B2567880  
LOCUS  
DEFINITION  
pac82-164\_7256.x1 pac82-164 Pseudomonas aeruginosa genomic clone  
pac82-164\_7256, genomic survey sequence.

ACCESSION  
B2567880  
VERSION  
B2567880.1 GI:27199936

KEYWORDS  
GSS.

ORGANISM  
Pseudomonas aeruginosa

SOURCE  
Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;

REFERENCE  
1 (Bases 1 to 1285)

AUTHORS  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

TITLE  
Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

COMMENT  
Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: Shotgun.

Location/Qualifiers

Source	1. 1285	/organism="Pseudomonas aeruginosa"	/mol_type="genomic DNA"	/strain="2-164"	/db_xref="caxon:287"	/clone="pacs2-164_7256"	/clone_1b="pacs2-164"	/note="clinical isolate 2-164 Whole genomic shotgun library"
BASE COUNT	219 a	405 c	355 g	289 t	17 others			
ORIGIN								
Query Match	9.1%	Score 115;	DB 29;	Length 1285;				
Best Local Similarity	53.3%	Pred. No. 1.6e-16;						
Matches 264;	Conservative	0;	Mismatches 230;	Indels 1;	Gaps 1;			
OY	670	GACGCTTGACAGCATCTGCTTCAAGGCGGACGCGGACGACGAGCTGAGCAGCAAGAGG	729					
DB	95	GACGCTTACAGCGGGCTGTGTGAGGCGCCCGACGAGCGGCGAGTTAGCGAAGCGGAA	154					
OY	730	CTGTGTGCGGCTGTGTGTGCGATTTATTCGCTGTGTGCGACCGATATCCAGCATCTTATC	789					
DB	155	CTGTCTTCATAGGCGCCACCTGTGTATGATGAGCGGCTTGTGAGCAGCATGATATATC	214					
OY	790	GCGTTGCTGTGTCTAACTGT	849					
DB	215	GCGAAGCGGCTGT	274					
OY	850	CCCGGGCTCATAGGAGGACGCGCTGTATGAGGTCGCGGCTTGCACATATCTCTGATATA	909					
DB	275	CCGGAACCTCTCTCCCAAGCCATGAGAAATCTGTGTGTGTGTGTGTGTGTGTGTGTGT	334					
OY	910	GGAACGTGTGCTTGT	969					
DB	335	TGATGT	394					
OY	970	GAGATGTCTTCTCTCTGTATCCCGGACGCGCTGTGAGATGTGGACCTGTATTTCTTCAAGGCA	1029					
DB	395	GAAATCATCTGT	454					
OY	1030	GACGT	1089					
DB	455	GACGCTCTGACCTCAGCCGCAACCGATGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT	514					
OY	1090	GTCGTCT	1148					
DB	515	TACTGCGT	574					
OY	1149	CCGTAGGTTCCCGCA	1163					
DB	575	CGCGGCTTCCCGCA	589					
RESULT 2								
LOCUS	BZ675691	975 bp	DNA	linear	GSS 05-FEB-2003			
DEFINITION	PUBDH78TD ZM 0.6 1.0 KB Zea mays genomic clone ZMBMT029M12,							
ACCESSION	BZ675691							
VERSION	BZ675691.1	GI:28226790						
KEYWORDS	GSS.							
SOURCE	Zea mays							
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
REFERENCE	Whitehead, C.A., Quackenbush, J., Van Aken, S., Utechtack, T., Resnick							
AUTHORS	Whitehead, C.A., Quackenbush, J., Van Aken, S., Utechtack, T., Resnick							
TITLE	Maize Genomics Consortium							
JOURNAL	Unpublished							
COMMENT	Contact: Cathy Whitehead							
	TIGR							

<b>FEATURES</b>						
source	location/Qualifiers					
	1..975 /organism="Zea mays" /mol_type="genomic DNA" /strain="873" /db_xref="taxon:4577" /clone="ZMMB1a029M12" /c1one_1lb="ZM 0.6-1.0 KB" /note="vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high Cor selected genomic DNA library"					
<b>BASE COUNT</b>	173 a	323 c	302 g	177 t		
<b>ORIGIN</b>						
Query Match	6.7%	Score 84.6;	DB 29;	Length 975;		
Best Local Similarity	49.3%;	Pred. No. 1.9e-09;				
Matches 250;	Conservative 0;	Mismatches 254;	Indels 3;	Gaps 1;		
Oy	668	ATAGACGCTTGACCATGCTGTTCACGGCCGACGCCGACAGCAGAGCGTGAAGC	727			
Db	216	ATTGACTTCATACCGCCCTTACCACCCGACGAAAGAAGCGAGTCCCTCAAGAGAC	275			
Oy	728	AGCTGTCGCGCTCGTGGTGCGATTATCGCTGTGCACCGATACCAAGATCACCTTA	787			
Db	276	AATTGCTGCTGTGTGATTTGCATCTCTGTGCGCGGTTAAGAACAGTGCCTCGACA	335			
Oy	788	TGCGGTTGCTGTGCTCAACTGTGTGCGGTGCCCGAGCGCTTCGAGCTGTGAAGCCG	847			
Db	336	TCCCCAATTATCCAGAGTGTCTGTGACCAACCCCGCAGTGGCAACAGCTCAAGCCG	395			
Oy	848	AGCCCGGAGCTCATAGGAACGCGCTCGATGAGGTGCTCCGCTTCGACATATCCTCAGAA	907			
Db	396	ATCCCGAGCAATACCGCAGCGGTCCAAAGCTGCTGCTCATATCCCGCTGGCTCGG	455			
Oy	.908	TAGAAA--CTGTGCGTTTTCGCCAGCAGGACCTGAAGTACTGCGGGCATTCATCAAGA	964			
Db	456	CGGGAATGTTTGTSCATACGCGCTGTGAACAATTCAGTCTGGTGAACCTCGTGCGCC	515			
Oy	965	AAGGGAGATGATGCTTTCTCTCTATATCCCGAGCGCCCTGAGAGATGGGACTGTATTCCA	1024			
Db	516	AGGCGAGCGCGGTGTGCTTCGATTCGCGCGCCGCAACATGACCCGCGCTTGAAA	575			
Oy	1025	GGCCAGACGTGTTGATGTGCGACGAGGACAACGCGCGGAGCCTTCGCTACGTAAGAGCC	1084			
Db	576	ACCACAGCGGCTGACCTGTGACGAGATGCAAGCGGACCATTTTCGGCTTTGGCCAGCGCC	635			
Oy	1085	CCCATGTCTGCCCCGGGGGTGTCTTTCCTGCTCGACGCGGAGATGCGCGTGGGACCA	1144			
Db	636	TGCATCACTGATGAGCTCAAGCCCTGCGCGGGGTGCAATTGCAAGAGAGCGCTGCAGCCC	695			
Oy	1145	TCTTCCTAGAGTTCCCGAGATGAAGC	1171			
Db	696	TGTGTGTGCGCTGCCAGCCTGTGAGC	722			
<b>RESULT 3</b>						
LOCUS	AY174775	330 bp	DNA	linear	GSS 13-JAN-2003	
DEFINITION	AY174775 Mycobacterium avium subsp. paratuberculosis DNA					
ACCESSION	AY174775					
VERSION	AY174775.1 GI:27689314					
SOURCE	GSS.					
ORGANISM	Mycobacterium avium subsp. paratuberculosis Mycobacterium avium subsp. paratuberculosis Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium					





**COMMENT**

Contact: Charles Hauser  
 DCMB Box 91000  
 Duke University  
 Durham, NC 27708-1000  
 Tel: 919 613 8159  
 Fax: 919 613 8177  
 Email: [chauser@duke.edu](mailto:chauser@duke.edu)

## FEATURES

## SOURCE

location/Qualifiers  
1. 609  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:1055"  
/clone\_lib="CC.reinhardtii CC-1690, Stress II (normalized  
, Lambda Zap II"  
/note="Vector: plasmid II SK-, Site\_1: EcoRI, Site\_2:  
XhoI, Stress condition II library, constructed by John  
Davies and Jeffrey McDermott, combines cDNAs from CC-1690  
cells grown to mid-log phase in TAP (NH<sub>4</sub><sup>+</sup> - containing)  
and shifted to TAP - NO<sub>3</sub><sup>-</sup> (24hrs); H<sub>2</sub> production  
conditions (0, 12hr, 24hr) see Meis et al., (2000) Plant  
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr): TAP +  
Pibitol (1, 2, 6, 24 hr), TAP + Cd (1, 2, 6, 24 hr).  
POLY A mRNA was purified from each sample, pooled and cDNA  
synthesized. The cDNA was directionally cloned into lambda  
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')  
sites. plasmid II SK- plasmids were excised from the  
lambda Zap clones by superinfection with ExAssist  
(Stratagene) phage. The library was normalized using  
method 4 described in Bonaldo et al., (1996) Genome  
Research 6: 791-806."

BASE COUNT	119 a	200 c	213 g	77 t
ORIGIN				

ORIGIN

Query Match	4.1%	Score 52;	DB 12;	Length 609;
-------------	------	-----------	--------	-------------

QY	568	CAGGTGATGAGAGACAAAGACCTGGTCGGTCCGTCACGAGAGGGAGCTCGCGTGC	627
Db	32	CAGCTACGGCGGGGTGGGCGCATGCGGGGTGGCACTGGGCTGGCGCGCTGGCCCCGGC	91
QY	628	CATGACGTCTCTGATGAGCGCGCGCAGGAACCCGCTCGAAAAATGACGTCTTGACGATGCTG	687
Db	92	CACCTGCTCTCTGTGACGAGCGCCCAACACCACTGGACATTTGTCACATCTGTGGCTG	151
QY	688	CTTCAGGGCGAGGGCCGAGCGGACGAGCTGAGCACTGAAAGAGCTGGTCCGGCTCTGGGT	747
Db	152	CAGAACCGGCTGGCCACTTGTGCGAGAGGGGAGAGGCTGGTGTGACGACAGACCGG	211
QY	748	GGCATTATGCTGCTGGCAACCGATACCAAGATCTTACCTTATCGGTGTCCGTGTGCTCAAC	807
Db	212	GGCTTCTCTCAACGGCGGTGACGAGGAGACATCATCTCTGGCGAGCGGGGCGCTGACCTAC	271
QY	808	CTGCTGCGGTGGCCCGGAGGCGCTCCAGCTGGTGGAAAGGCGGAGCC	851
Db	272	TTCGCCGCGCGCTTACGACCGCTTACATGACGAGGGAAGGCGGAGGCG	315

RESULT 8				
CNS0091P/c				
LOCUS	CNS0091P	925 bp	DNA	linear
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # GSS 03-JUN-199			

ACCESSION	AL053013
VERSION	AL053013.1
KEYWORDS	GI:4934461
SOURCE	GSS.
ORGANISM	Drosophila melanogaster (fruit fly)
	Drosophila melanogaster

REFERENCE	1 (bases 1 to 925)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1997)

BP 191 91006EVRY cedex - FRANCE (e-mail : segetre@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this Barcode sequence was carried out as part of a collaboration with the Bartley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oseegawa and Aaron Mamoser in Pierer de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BDGP Resource Center can be found at [http://bdgpac.med.buffalo.edu/drosophila\\_bac.htm](http://bdgpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES**  
**SOURCE**

**Source**

```

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_1b="RPCI-98"
/notes="end : TET3"

```

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN					

ORIGIN

Query Match	4.1%	Score 51.8	DB 29	Length 925
Best Local Similarity	14.5%	Pred. No. 0.092		
Matches 52	Conservative 159	Mismatches 148	Indels 0	Gaps 0

[illegible]

RESULT	9			
LOCUS	BG488189			
DEFINITION	BG488189	425 bp	mRNA	linear
	RHIZ2.60.F03.b1.A003			
	Rhizome2 (RHIZ2)			
	Sorghum propinquum cDNA, mRNA			EST 27-MAR-2001

ACCESSION	BG488189
VERSION	BG488189.1
KEYWORDS	GI:J3469424
SOURCE	EST.
ORGANISM	<i>Scorhium propinquum</i>
	<i>Scorhium propinquum</i>
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatocyt.; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD







REFERENCE	1 (bases 1 to 596)
AUTHORS	Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
TITLE	USA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation
JOURNAL	unpublished
COMMENT	Contact: Tyler B

1860 Pratt Dr., Blacksburg, VA 24061, USA  
Tel.: 540-231-7318  
Email: bmtylet@vt.edu  
PCR Primers  
FORWARD: BK reverse  
Plate: 008 row: I column: 18  
Seq primer: BK reverse  
High quality sequence stop: 596.  
I will be in the field for 10 days.

FEATURES	Location/Qualifiers
source	1. .596

```

/Note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      119 a      190 c      212 g      75 t
ORIGIN

```

Query Match	4.0%	Score 50.2	DB 14	length 596
Best Local Similarity	45.5%	Pred. No. 0.19		
Matches 178; Conservative	0	Mismatches 213	Indels 0	Gaps 0

OY	546	GCTGGCGCTGGGGTTTGGTGGCCCAAGTGGATAGAGGACCAAGACCTCGTGGCGGCGGT	605
Db	67	GCGCGTGGTGGACTACGCGCGGTCAAGTTCGGCTGGAGCCCAAGGGGGTGGACTTCAAGA	126
OY	606	CACCGAGGGGGCTCGCGCTGCTCCATGACGTCCTCGATGAGCGCGCGAGGACCCTGGTGA	665
Db	127	CGTCAAGATGAGCATGAAACCCCTTCTGGAGATTCGCGGTGGAGGAGGCCATCGGCTTCA	186
OY	666	AAATGACGCTTTGATGATGATGCTGCTTCAAGCGCGAGGCCGACGACGAGCTGACACGA	725
Db	187	GGAGAAAGACTGGCGCACCGAGGTGATGGCCCTGTCTCATCGGCCCCCAAGCAGCGAGA	246
OY	726	GGAGTGTATCGGCGCTCGTGGGTGGCATTATCGCTGCTGGACCGATACCAGATCTACCT	785
Db	247	GACGCTGGCGACAGGGCGCTGGCCCATGAGCGCGGACCGCGCATTCACATACACACGGACAT	306
OY	786	TATCGCGTTTCGCTGTGCTCAACCTGCTCGCGGTGCGCCGAGGCGGCTCGAGTGGTGAAGC	845
Db	307	GCGCACGGACCAAGAGAGTGCAGCCCGCTGGCCGTGGCCAAAGCTGCTCAAGAGAGTCTGTGC	366
OY	846	CGAGCCCGGGCTCATGAGGAAACGCGCTCGATAGGTGCTCGGCTTTCGACATATATCTCAG	905
Db	367	CAAGGAGAGCGCGACGCTCGTCATCTGCGGCAAGACAGACATCGACGCCAGCGCGCA	426
OY	906	AATAGGAATCTGTGCGTTTTCGACGACGACGAC	936
Db	427	GACGGGCCCATGCTGGCCGGGCTTCTTGAC	457

RESULT	14
CD043612	
LOCUS	
DEFINITION	CD043612 610 bp mRNA linear EST 09-MAY-2003 psHB009XH01F.180418 psHB: Infected hypocotyl soybean host. 48 hrs post infection Phytophthora sojae cDNA clone SHB009H01 5' mRNA sequence.
ACCESSION	CD043612
VERSION	CD043612.1
KEYWORDS	GI:30497205 EST.

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Phytophthora sojae	Phytophthora sojae	1 (bases 1 to 610)	Tyler, B.M., Judelson, H.S., Gijzen, M., Dean, R.A. and Vaughn, M.E.	USDA-FFARS: Expression of Phytophthora sojae genes during infection and propagation	unpublished	Contact: Tyler B

1880 Pratt Dr., Blacksburg, VA 24061, USA  
Tel: 540-231-7338  
Email: bmcj1er@vt.edu  
PCR Primers  
FORWARD: BK reverse  
Plate: 009 row: H column: 01  
Seq primer: BK reverse  
High quality sequence spots: 610.

FEATURES	Location/Qualifiers
source	1. .610

```

clone_1b="pSHB: Infected hypocotyl soybean host. 48 hr
post_infection"
note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI;
USDA-IPAPs: Expression of Phytophthora sojae genes during
infection and propagation."
BASE COUNT      121 a      199 c      215 g      75 t
ORIGIN

```

Query Match	Similarity	4.0%*	Score 50.2	DB 14	Length 610
Best Local	Similarity	45.5%*	Pred. No. 0.13		
Matches 178	Conservative	0	Mismatches 213	Indels 0	Gaps 0
Qy	546	GCTGCGCGGTGGTTGTGTGCCAGGTGATGAGAGACCAAGACCTGTGCGCTCGT	605		
Db	69	GCGGGTGGTGACTACGCCGCTCAAGATCCGCTGAGGCCCAAGGCGTGGACCTTCAMANA	128		
Qy	606	CACCGAGGGGTCCCGCTGCTCCATGACGTCTGTGATGAGGGGGCAGAAACCTGGCTGA	665		
Db	129	CGTCAAGTATGCACTGAACCCCTTCTGTGAGATGCCCTGGAGGAGGCCATTCGGCTCAA	188		
Qy	666	AAATGACGTTCTGACGATGCTGCTTCAAGGCCGAGGCCGACGGCAGCAGGCTTGAGCAGAA	725		
Db	189	GGAGAGAGACTGGCCACCGAGGTCGTGGCGCTGTCCATCGGCCCAAGCAGAGCCAGGA	248		
Qy	726	GGAGCTGTCCCGCTCGTGGGTGGATTATGCTGTGTGGCACCCATACACAGATTTACTT	785		
Db	249	GACCTGTGGCAGCGCGCTGGCCATGGGGCGGACCGCGGCATTCACATACCAACGACAT	308		
Qy	786	TATGCGCTTCTGTGCTCAACTGCTGTGCGGTGCCGCCAGGCGCTGCAGCTGTGTAAGAC	845		
Db	309	GCGCACGAGCCAGAGAGCTGCACGCCGCTGGCCGTGCCAAGCTGTCTAAGGAAGTGTGTGGC	368		
Qy	846	CGAGCCCGGGCTCATGAGAAACGGCTGATGAGGTGCTCCGCTTGCACATATCTTCAG	905		
Db	369	CAAGGAGAGCCGACGCTGCTCATCTGGCGGCAAGACAGACATGACGCCGACGCCGCA	428		
Qy	906	AATGAAACTGTGGCTTTCGCGCAGGACGAC	936		
Db	429	GACGGGCCCATGCTGGCGCGGCTGTGTGGAC	459		

RESULT 15			
CD262676			
LOCUS			
CD262676	623 bp	mRNA	linear
			EST 23-MAY-2003

DEFINITION pSMA018K21f.191994 sMA: Phytophthora sojae grown in synthetic medium  
ACCESSION CD262676  
VERSION Phytophthora sojae cDNA clone sMA018K21 5, mRNA sequence.  
KEYWORDS EST.  
SOURCE Phytophthora sojae  
ORGANISM Phytophthora sojae  
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
Phytophthora.  
REFERENCE 1 (bases 1 to 623)  
AUTHORS Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.  
TITLE USDA-IRAFs: Expression of Phytophthora sojae genes during infection and propagation  
JOURNAL Unpublished  
COMMENT Contact: Tyler B  
Tyler lab  
VBI  
1880 Pratt Dr., Blacksburg, VA 24061, USA  
Tel: 540-231-7318  
Email: bmtyle@vt.edu  
PCR Primers  
FORWARD: BK reverse  
Place: 018 row: K column: 21  
Seq primer: BK reverse  
High quality sequence stop: 623.  
Location/Qualifiers  
1..623  
/organism="Phytophthora sojae"  
/mol\_type="mRNA"  
/db\_xref="taxon:67593"  
/clone="sMA018K21"  
/tissue\_type="mycelium"  
/cell\_line="P6497"  
/dev\_stage="mycelium"  
/clone\_1fb="sMA: Phytophthora sojae grown in synthetic medium"  
/note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"  
BASE COUNT 123 a 206 c 217 g 77 t  
ORIGIN  
Query Match 4.0%; Score 50.2; DB 14; Length 623;  
Best Local Similarity 45.5%; Pred. No. 0.19;  
Matches 178; Conservative 0; Mismatches 213; Indels 0; Gaps 0;  
QY 546 GGTGGCGGTGGTGGTGGCCAGGTGATGAGAGACCAAGACCTGGTGGTGGT 605  
DB 59 GGGCTGTGGACTACGCCCTCAAGATCCGCTGAGGCCAAGGGGTGACCTCAAGAA 118  
QY 606 CACCGAGGGGCTCGGCGTGTCTCATGACGTCTCGATGAGCGCGCAGGAACTCCGTCGA 665  
DB 119 CGTCAAGATGACATGAACCCCTTCTGCGAATCCCGTGGAGAGGCCATCCGCTCA 178  
QY 666 AAATGACGCTTGTACGATGCTGCTTCAAGCGGAGGCCGAGCGGACAGGCTGAGCAGAA 725  
DB 179 GGAGAGAAAGCTGGCCACCGAGGTGCTGGCCGTGTCATCGGCCCCCAAGCAGACAGGA 238  
QY 726 GGAGCTGTGGGCTCGGCGTGTGCGATATATCGCTGTGCAACCATACGATCTACT 785  
DB 239 GACGCTGGCGACGGCGCTGGCCATGAGCGCGGACCGGATTCATCAACCGACAT 298  
QY 786 TATCCGCTTGTGCTGCTCAACCTGCTGGGTGCGCCGAGCGGCTGAGCTGTGAAGGC 845  
DB 299 GGGCACGACGACGAGAGCTGCAACCGCTGGCCGCTGACCAAGCTGCTCAAGAGAGTGTGGC 358  
QY 846 CGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTCGCTTGCACATATCTCAG 905  
DB 359 CAAGAGAGGCGCGAGCTGCTCATCTGCGCAAGCAGAGCATCGACGCCGACCGCGCA 418  
QY 906 AATGAACTGTGCTTGGCCAGGAGGAC 936  
DB 419 GACGGCCCTCATGCTGGCGGCTGCTGAC 449

Search completed: October 3, 2003, 17:48:11  
Job time : 2496.11 secs

**THIS PAGE BLANK (USPTO)**



Db 507 -----GGAGCGCCAGCGCGGAGACTCTGTCCCGGAGGAGCCAGCCGCTGCGCGATG 559  
Qy 614 GGCTGCGCTGCTGCATGACGTCCTTCGATGAGCGGCGAGGAAACCCGCTCGAAAAATGAG 673  
Db 560 CCGAGGTGACATACCTGCGCTGTCTGTCTGAGAGCCAGAGCGCGGCGAGCGAGCG 619  
Qy 674 TCTTGACGATGTCTGCTTCAAGCGCCGAGCCGAGCGGAGGCTGAGCAAGAGAGCTGG 733  
Db 620 TCTTACGCGCGGCTGTGTGAGCGCGCGCGAGAGCGCGGAGGCTGAGCAAGAGGAACTGG 679  
Qy 734 TCGCGCTGCTGAGTGTGATTCGCTGTGTCGACCGATACCAAGATCTTACTTATGCGCT 793  
Db 680 TCTTCATGAGCCACCTGCTGATGATGAGCGGCTTGTGAGACCAATGAAATGATGCGA 739  
Qy 794 TCGCTGTGCTCAACCTGCTGCGGTGCGCGGAGCGCTGAGCTGTGAAAGCGGAGCGCG 853  
Db 740 AGCGGCTGTGACCTGCTGTGTCAACCCGAGAGCACTGGCGTGTGCGGCGAGCGCG 799  
Qy 854 GGCTCATGAGAAAGCGGCTGATGAGTGTCTCGCTTTCGACAAATCTCAGAAATGAGA 913  
Db 800 AACTCTGTGCCAAGCGCATGAGAAAGTGTGTCCGACAGATCGGCTGCGGCTTCA 859  
Qy 914 CTGTGCGTTCGCGAGCGAGACCTGAGTACTGCGGCGCATGATCAAGAAAGGAGAGA 973  
Db 860 TGTGTGCTTCAACCGTGAAGAGTGAAGTGAAGCGGCTCACTTCCCGCGGAGAT 919  
Qy 974 TGTCTTCTCTGATCCGAGCGCTTGAAGATGGAATGTATTTCTCAGGCGAGAG 1033  
Db 920 ACATCTGTGCTTCAACCTGAGCGGAGCGCAACAGATCCGAGCGCTTCAAGATCCGAG 979  
Qy 1034 TGTGTGATGTGCGAGCGGAGCGGCGGAGCGCTGCGGATGAGTGAAGGCGCGCGATGCT 1093  
Db 980 GCGTGTGCTTCAACCGGAGCGGAGCGGATGCGCATCTGTGCTTGTGCGGCTGCTACT 1039  
Qy 1094 GCGCGCGGAGTGTCTGCTGCTGCGCTGAGCGGAGATCGCGGAGGAGCAATCTTCCGTA 1153  
Db 1040 GCGTGTGCGCTGTGCTGCGCGGCTGAGAGGCGGAGTCCCATCCAGCGGCTGTGCGCG 1099  
Qy 1154 GGTTCGCGAGATGAAGCTGAAAGAACTCCGCTG 1188  
Db 1100 GCTTCCCGACCTTCAGTGTGCGGCTGCCCGAGCG 1134

RESULT 2  
PCT-US03-18787-2/c  
Sequence 2, Application PC/TUS0318787  
GENERAL INFORMATION:  
APPLICANT: KOSAN BIOSCIENCES  
APPLICANT: Hutchinson, Richard C.  
APPLICANT: Reid, Ralph C.  
APPLICANT: Hu, Zhihao  
APPLICANT: Rascher, Andreas  
APPLICANT: Schirmer, Andreas  
APPLICANT: Medaniel, Robert  
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
TITLE OF INVENTION: PRO-GLUTAMININ PRODUCING POLYPEPTIDE SYNTHASES AND  
FILE REFERENCE: 30622009740  
CURRENT APPLICATION NUMBER: PCT/US03/18787  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/389,255  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: US 60/393,929  
PRIOR FILING DATE: 2002-07-03  
PRIOR APPLICATION NUMBER: US 60/395,275  
PRIOR FILING DATE: 2002-07-12  
PRIOR APPLICATION NUMBER: US 10/212,962  
PRIOR FILING DATE: 2002-08-05  
PRIOR APPLICATION NUMBER: US 60/415,326  
PRIOR FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: US 60/420,820  
PRIOR FILING DATE: 2002-10-24  
PRIOR APPLICATION NUMBER: US 60/433,130

Qy 303 GGATACGCTGGGGTCCGCAAGCTGTGTAACCCGCTGTTTACGTACGCGCCGATGACT 362  
Db 82492 GGAGCAGCCCGGAGTCCCGCATGCTACCGCGAACTTCAACATCCGCGCACCGAGGC 82433  
Qy 363 GGTGCGCGCCGAATACAGCGGACCGTGCAGACGCTGCTGATGCTGCTCGGACAGA 422  
Db 82432 GATGCGCGCGATGATCCAGGCGATGTCAGAGGCTCTTGAACCGGCTGATGCGCGAG 82373  
Qy 423 ---GGAGTTGACGTTGTGCGGATTTACGCGAGGAAATCCGATGCGCGGATCAGCGC 479  
Db 82372 CCGCGCGCGGACCTGTGCGGAGTTCGCTTCCCTGCGGCTGAGTGTGCGGCT 82313  
Qy 480 TCTGTGAAAGTTCGCGCGGAGTGAAGAGTTCGCTGCTGCTGCGGCTGCGGAGTGC 539  
Db 82312 GATGCTGCGGCGTTCGAGCGCGGACTTGTGCGGAGTTCCAGAG-----GC 82268  
Qy 540 GCGCGCGCTGCGGCGTGTGTTGTGTCGCCAGGTCATGAGAGCAAGACCTGCTGCG 599  
Db 82267 GTTCGAGGCGCTGATGACTTCAACCGCTGCGCGGAGAGATGGCGCGGCTGCGGCT 82208  
Qy 600 GTTCGTCACGAGGCGCTGCGCTGCTCATGACGCTCTGATGAGCGGCGGAGAACCC 659  
Db 82207 CATGTGAGATGAGGAGCGGAGTGTGCGGCGGAGCGCGGAGCGGCGGAGATCT 82148  
Qy 660 GGTGCAAAATGACGCTTGAAGATGCTGCTTCAAGCGGAGCGGCGGAGCGGAGCTAG 719  
Db 82147 GCTCA-----GCGGCTCATGTGTGACCAAGAGCGGAG-----GCGGCTCAC 82103  
Qy 720 CACGAAGAGCTGTGCGGCTGTGCGGATTTATCGCTGCTGCGGAGCGGATGATGAT 779  
Db 82102 CAGAGAGAGTGTGTGCGGAGCGGCTGTGCTGCTGCGGAGCGGAGCGGAGCGGAG 82043  
Qy 780 CTACCTTATGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839  
Db 82042 CAACATGATGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81983  
Qy 840 GAAAGCGGAGCGGCGCTGATGAGAGAGCGGCTGATGAGGCTGCTGCTGCTGCTGCTGCT 899  
Db 81982 GCGGCGCGAGTGTGCGGCTGATGAGAGAGCGGCTGATGAGGCTGCTGCTGCTGCTGCT 81923  
Qy 900 CCTCAGAAATGAGAACTGTGCGGCTGCGGAGGAGCTGAGTACTGCGGCGGCTGCTGCT 959  
Db 81922 CCTCAGAAATGAGAACTGTGCGGCTGCGGAGGAGCTGAGTACTGCGGCGGCTGCTGCT 81863  
Qy 960 CAAGAAAGGAGATGTCTTCTCTTATCCGAGCGGCTGAGAGATGAGACTGTATT 1019  
Db 81862 CCGGCGCGGAGAGGAGTGTATCATCACTGCGGCGGAGCGGAGCGGAGCGGAGCGG 81803  
Qy 1020 CTCAGAGCGGAGGCTGTGATGATGAGGAGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1079  
Db 81802 GAGCGCGGAGGCTGTGAGCTGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 81743  
Qy 1080 AGGCGCGGAGTGTGCGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139  
Db 81742 CCGGCGGAGCGGAGTGTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 81683  
Qy 1140 CACCATCTTCTGAGGCTTCCGAGATGAGTGAAGAACTCCGCTG 1188  
Db 81682 GACCTGCTGCGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 81634

RESULT 3  
PCT-US03-19069-2/c  
Sequence 2, Application PC/TUS0319069  
GENERAL INFORMATION:  
APPLICANT: KOSAN BIOSCIENCES  
APPLICANT: Hutchinson, Richard C.  
APPLICANT: Reid, Ralph C.  
APPLICANT: Hu, Zhihao  
APPLICANT: Raecher, Andreas  
APPLICANT: Schlimer, Andreas  
APPLICANT: McDaniel, Robert  
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
TITLE OF INVENTION: PRO-GEIDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
FILE REFERENCE: 300622009740  
CURRENT APPLICATION NUMBER: PCT/US03/19069  
CURRENT FILING DATE: 2003-06-16  
PRIOR APPLICATION NUMBER: US 60/389,255  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: US 60/393,929  
PRIOR FILING DATE: 2002-07-03  
PRIOR APPLICATION NUMBER: US 60/395,275  
PRIOR FILING DATE: 2002-07-12  
PRIOR APPLICATION NUMBER: US 10/212,962  
PRIOR FILING DATE: 2002-08-05  
PRIOR APPLICATION NUMBER: US 60/415,326  
PRIOR FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: US 60/420,820  
PRIOR FILING DATE: 2002-10-24  
PRIOR APPLICATION NUMBER: US 60/433,130  
NUMBER OF SEQ ID NOS: 153  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 86941  
TYPE: DNA  
ORGANISM: Streptomyces hygroscopicus  
PCT-US03-19069-2

Query Match 10.3%; Score 130.2; DB 1; Length 86941;  
Best Local Similarity 50.4%; Pred. No. 1,1e-24;  
Matches 448; Conservative 0; Mismatches 408; Indels 33; Gaps 4;

QY 303 GGATCACCCTGGGCTCCGCAAGCTGCTCAACCCGCTGTTAGTCAGCGCCATGCACT 362  
DB 82432 GGAGACCCGCGGATCCCGCGATGCTCAACCGCAATTCACATCCGCGCAGGAGC 82433  
QY 363 GCTGCGCCGCAATATACAGCCGACCGTGCATGCTGCTCCGCGCAAGA 422  
DB 82432 GATGCGCGCATGATCCAGGGGATGTCGACGGGCTCTGGAACGGCTGATGCGCCAGG 82373  
QY 423 ---GAGATTCCAGCTTGTGCGGGAATTAGCGGAGGAATCCGATCCGCGCATAGGC 479  
DB 82372 CCGCGCGCGCACTGCTGCGGCGGATCTTCCCGCTCCGCGTCCAGGTGATGCGGCT 82313  
QY 480 TCTGTTGAAGTTCGCGCGCAAGTGAAGAAAGTTCCGCTGCGCTCGGCTGCGGACTGC 539  
DB 82312 GATGCTGGGGGCTTCTGGAACCGCACTTGGCGAAGTTCCAGG-----GC 82268  
QY 540 GCGCGCGCTCGCGGCTGCTGCTGCGCGGCTGATGAGAGCAAGACCCCTGATGCGC 599  
DB 82267 GTCGAGGGGCGCTGATGAGCTTACCGCGCTGCGCGAGAGAGGCGCGGCTGCGGCT 82208  
QY 600 GTCCGCTACCGAGGGGCTGCGCTGCTGATGAGCTCTTCAATGAGCGCGCGAGAAACC 659  
DB 82207 CATGTGAGCTACGAGCCCGATGTCGCGCGCAAGCGCGCGACCGCGCGAGATCT 82148  
QY 660 GCTCAAAATGACGCTTGAAGATGCTGCTTCAAGCCGAGCGGAGAGGCTGAG 719  
DB 82147 GCTCA-----GCGGCTCATGCTGATGACAGAGCCGAGCG--GCGGCTCAC 82103  
QY 720 CACGAAGAGCTGTGCGGCTGCTGCGGATTAATGCTGCTGCGACCGATACAGCAT 779

DB 82102 CCAGACGAGCTGTGCGCACCGCCCTGCTGCTGCTGCGCGCACAGACCAACCGC 82043  
QY 780 CTACCTTATGCGGCTTGTGCTGCTCAACCTGCTGCGCGCGCGCGCTGAGCTGCT 839  
DB 82042 CAACATGATCCCGCTGCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81983  
QY 840 GAAGCCGAGCCGCGGCTCATAGAGAAAGCCGCTGATGAGTGTCTCCGCTTCAATAT 899  
DB 81982 GCGGCGGATGCTGCGGCTGATGAGGCAAGCGGTGAGCAATGCTCCGCTGATACAT 81923  
QY 900 CTTGAAATGAGAACTGTGCTTTCCTGATGCTCCGAGCGCCCTGAGAGATGAGACTGTATT 959  
DB 81922 GCTCAGAGAGGCAACCGGAGCGGTGCTGCTGAGAGACTGAGAGTGGCGGTACTCAT 81863  
QY 960 CAAGAAAGGAGATGCTTCTTCTGATGCTCCGAGCGCCCTGAGAGATGAGACTGTATT 1019  
DB 81862 CCGGCGGATGAGGAGGATGATCATTAACCTGCTCCAGCGCCCAACCGGAGCCCACTTGC 81803  
QY 1020 CTCCAGGCGCAGACGCTGTTGATGTCGACGAGAGACGCGCGCGAGCTGCGTACGATG 1079  
DB 81802 GAGCGCCACGAGCTGACCTGAGAGCGCGCCCAAGCGCGGAGATGTCGGTTCGCTT 81743  
QY 1080 AGGCCCCATGCTGCTCCCGGCGGTGCTCTTCTGCTGCTGCTGAGGCGAGATGCGCTG 1139  
DB 81742 CGGCGTGACAGGCTGCTGCGGCGAGACCTGCGCGGCTGAGCTCCAGATGCTCTGGA 81683  
QY 1140 CACCATCTTGGTGTGCTTCCCGGAGATGAGTAAAGAACTCCCGG 1188  
DB 81682 GACCTGCTGCGCGCGCTGCGCCAGCTGCGCTGAGGTCCCGCTCGAG 81634

RESULT 4  
PCT-US03-18787-1/c  
Sequence 1, Application PC/TUS0318787  
GENERAL INFORMATION:  
APPLICANT: KOSAN BIOSCIENCES  
APPLICANT: Hutchinson, Richard C.  
APPLICANT: Reid, Ralph C.  
APPLICANT: Hu, Zhihao  
APPLICANT: Raecher, Andreas  
APPLICANT: Schlimer, Andreas  
APPLICANT: McDaniel, Robert  
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING  
TITLE OF INVENTION: PRO-GEIDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND  
FILE REFERENCE: 300622009740  
CURRENT APPLICATION NUMBER: PCT/US03/18787  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/389,255  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: US 60/393,929  
PRIOR FILING DATE: 2002-07-03  
PRIOR APPLICATION NUMBER: US 60/395,275  
PRIOR FILING DATE: 2002-07-12  
PRIOR APPLICATION NUMBER: US 10/212,962  
PRIOR FILING DATE: 2002-08-05  
PRIOR APPLICATION NUMBER: US 60/415,326  
PRIOR FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: US 60/420,820  
PRIOR FILING DATE: 2002-10-24  
PRIOR APPLICATION NUMBER: US 60/433,130  
NUMBER OF SEQ ID NOS: 153  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 85692  
TYPE: DNA  
ORGANISM: Streptomyces hygroscopicus  
PCT-US03-18787-1

Query Match 9.9%; Score 124.6; DB 1; Length 85692;  
Best Local Similarity 48.0%; Pred. No. 3.5e-23;





```

Qy 553 GTGGTTTGTCGCCAGCTGATGAGAACCAACCTGTCGCTCCGTACCCGAG 612
Db 3746 CCG-----ACCGACCCCGGACCGCCGCTGTATGAGAGCCCGCTGCAC 3705
Qy 613 GGGCTCGGCGTCTTCATGACGTCCTCGATGAGGGGGCCGAGAACCCGCTGAAATAC 672
Db 3704 GGGCTCCAGACATATAGCCGAGGTCGTG---CCCATAGCCGGAACACGGGAGGAC 3648
Qy 673 GTCTTGACGATGCTGCTTCAAGCCGAGGCGCAGCGACAGCTGAGCAGCAAGAGCTG 732
Db 3647 GACCTGCTCAGCCGCTGCTGTGTGCGCCGACCGCAGCAAGAACTCTCCAGAGAGCTG 3588
Qy 733 GTCCGCTGCTGAGGTCGATTAATGCTGCTGCGACCGATACCAAGATTAATGCGG 792
Db 3587 GACTCGATGATCTTCCAGCTCCTCGTGGCGGGCCAGAAACGGTACCAACAGATCAC 3528
Qy 793 TTCCGCTGCTCAACCTGCTGCTGCTGCGCCGAGGCGCTGAGCTGTGAAGCCGAGCC 852
Db 3527 ACCGCTGATGCTGCTTCTTCCGACCCCGCAACTGCGCCGCTGCGGACGACCCG 3468
Qy 853 GAGCTCATGAGGAACGCGCTGATGAGGTCCTCCGCTTCAACAATATCTCAAGATAGA 912
Db 3467 GAGCTGATGCCCCGCGCGCTGACGAACTCTCCGCTACGACAGCGCTTGAAGTAC 3408
Qy 913 ACTGTGCTTTGCGCAGGACCTGAGATCTGCGGGGCAATGATCAAGAAAGGGAG 972
Db 3407 ACTGCGCCCTTCTTTCGACAGGACAGCACTGACGCGACGAGCTCCGCGCGGAG 3348
Qy 973 ATGCTCTTCTTCGATCCCGAGCGCCTGAGAGATGAGGATCTTCTTCAAGCGCAAC 1032
Db 3347 TCGGATGATCTTCTTCTTCTGCGCGCAACGCGCAACCGCGCTTCCGACCCAC 3288
Qy 1033 GTGTTGATGTCGACGAGGACAGCGGCGCGAGCTCGGTAAGAGAGAGCCCATGTC 1092
Db 3287 ACCCTGACCTGACCGCTGACCCCAACCCCACTGGCTTGGTCAAGGATCACTTC 3228
Qy 1093 TGCCCCGGGCTGCTTCTGCTGCTGCGCTGAGGCGGAGATGCGGTGGGACCAATCTTCGT 1152
Db 3227 TGCCCCGGGCGCGCGCTGCGCGCGCAACTCCAGATGCGGTGGGACCACTCTGTC 3168
Qy 1153 AGCTTCCCCGAGATGAACTG 1173
Db 3167 CGCTCTCCCGAGCGCACCTG 3147

```

RESULT 6  
US-10-297-465B-1/c

```

; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Secubal, Joao
; APPLICANT: Medanis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; PRIOR FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

```

Query Match 5.4%; Score 68.6; DB 6; Length 2731748;  
Best Local Similarity 46.2%; Pred. No. 1.4e-07;  
Matches 264; Conservative 0; Mismatches 304; Indels 3; Gaps 1;

```

Qy 602 CCGTACAGAGGGCTGCGCTGCTCATGACGTCCTGATGAGCGGCGCAGAACCCGC 661
Db 38855 CTGCTTACGAAGAGCTTCCGCAATATCTTACGAATGATGATGACTACGCCGACCATC 388496
Qy 662 TCGAAATGAGCTTTTACGATGCTGCTTCAAGCGCGGAGCCGACAGCGAGCTGAGCA 721
Db 388495 CTGGAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 388436
Qy 722 CGAAGAGCTGCTGCGCGCTGCGGCTGCGATTAATGCTGCTGCGACCGATCAAGATCT 781
Db 388435 ATGATGAATGCTTCTCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 388376
Qy 782 ACCTTATGCGCTGCTGCTGCTCAACCTGCTGCGGCTGCGCGCGCTGAGCTGTA 841
Db 388375 ATATGATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 388316
Qy 842 AGCCCGAGCCCGGCTCATGAGAACCGCTGATGAGTGTCTCCGCTTCAACAATATCC 901
Db 388315 AGAGTGAATCTGTCACTGATGCAACAGCGGATGCGAATGCTTAAGTCAAGCGGTCGG 388256
Qy 902 TCAGATAGAGAACTGCGCTTTCGCGAGGACGCTGAGTACTGCGGGGCTCATGATCA 961
Db 388255 TACAGTTC---ACGATGCGCGCGCGATGATGATTAATGAGGTTGAAGCGAGTGTGCTC 388199
Qy 962 AGAAGGAGAGATGCTTCTTCTCTGATCCGAGCGCTGAGAGATGGAATGATATCT 1021
Db 388198 CTGCTGAGCAGTATGATATTTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 388139
Qy 1022 CCAGCCAGAGCTTTGATGATGCTGACAGGACAGCGGCGGAGCTGCGCTGATGAG 1081
Db 388138 CTCAATCCCGACCACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 388079
Qy 1082 GCGCCAGATGCTGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141
Db 38078 GCATTCATCATGCTTGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 388019
Qy 1142 CCATCTTCGATGAGTTCGCCGAGATGAGCT 1172
Db 388018 CGCTTTGAACGCTTACCGCATTTAAGGCT 387988

```

RESULT 7  
US-10-297-465B-1

```

; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Secubal, Joao
; APPLICANT: Medanis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; PRIOR FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

```

Query Match 4.4%; Score 55.8; DB 6; Length 2731748;  
Best Local Similarity 45.4%; Pred. No. 0.0004;  
Matches 241; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

Dd	363970	AAGCTGCGCCGGCGCCAGCCTGGAATGCACTGATTTCATGTTGCTGCGCTGAGAGAG	364029	
Qy		703	GACGCGACAGGCTGTAGACACGAAAGAGCTGTCGCGCTGTGGGTGCGATTATTCGCTGCT	762
Dd	364030	GAGGGCCAGAAACTGACCGCATGATGAAATCGTGTCCAAATGTGATTTTGTGTTATTTGGGT	364089	
Qy		763	GGCACCGCATACCAACGATTCACCTTATTCGGCTTCGTGCTCAACCTGCTGCGGTCGCC	822
Dd	364090	GGCTACGAGACCAACATCCAAATCGAATTTGGCAATGCAATTGATTTGCTTGCATTCGTCAATCG	364149	
Qy		823	GAGCGCGCTCGACCTGTGTGAAGCCGACCCGGGCTCATAGGAACGCGCTCGATGAGGTG	882
Dd	364150	AAGCAGCTTGCACGACTCAAGAGTGTATTTGCTGCTATGCGCGCAACGAAATTTGGAATGC	364209	
Qy		883	CTCCGCTTCGACAAATATCTCGAAGAAATGGAACCTGTCGTTTTGCCAGGCGACGACTGAG	942
Dd	364210	TTACCTTACAGATGATCGGTACAGTTCC--ACGATCGCGCGCGCATGATGATGTGAGC	364266	
Qy		943	TACTGCGGGGCATCGATCAACGAAGGAGATGTCCTTCTCCTGATCCCGACGCGCTTG	1002
Dd	364267	ATTGAAGGTGATGTAGTACCTCGTGGCAGATTTGTGTTTGTGATGCTTGGTGTGCTTAAC	364326	
Qy		1003	AGAGATGGGACTGTATTTCTCCAGGCGCACGTCGTTGATGTGCGAGGGGACAGCGGCGG	1062
Dd	364327	CGTATCCGGCGCCCAATTCACCTGATCCGATCATCTGACAATCAACAGGAAGCAAGAGCGC	364386	
Qy		1063	AGCTTCGCGTACGATAGAGAGCCGCCCATGTCGCGCGGGGATGTCCTTGTGCTGCGCTGAG	1122
Dd	364387	CTCCAGTCAATTGGCGCGAGGCGTTCATCATGTGCTTGATATGCACTTGGGTGGTTGAG	364446	
Qy		1123	GCGGAGATCGCCGTGGGACCACTTTCCTGAGTTCCCGAGATGAAGCTG	1173
Dd	364447	TTGGATGCGCATTCGCTGTGTTGGAGCGCTCCGCAATTGAGGTTG	364497	
RESULT 8				
US-60-501-821-38				
; Sequence 38, Application US/60501821				
; GENERAL INFORMATION:				
; APPLICANT: ECOPIA BIOSCIENCES INC.				
; APPLICANT: Mcaldine, James				
; TITLE OF INVENTION: POLYENE POLYKETIDES AND METHODS OF PRODUCTION				
; FILE REFERENCE: 3010-1P				
; CURRENT APPLICATION NUMBER: US/60/501,821				
; CURRENT FILING DATE: 2003-09-11				
; NUMBER OF SEQ ID NOS: 57				
; SOFTWARE: PatentIn version 3.0				
; SEQ ID NO: 38				
; LENGTH: 11505				
; TYPE: DNA				
; ORGANISM: Amycolatopsis orientalis				
US-60-501-821-38				

	Query Match	4.3%	Score 54	DB 7	Length 11505
	Best Local Similarity	45.0%	Pred. NO. 0.00021		
	Matches 244	Conservative 0	Mismatches 299	Indels 3	Gaps 1
QY	278	ACGATTGTCGGGCTGC	CGCCGCGAGATCA	CGCTCGGATCGCCAGCTCGTCAACCCGT	337
Db	9509	ACCGGATCTTCGGCCTGTT	CGACGGCGGCTTCGAC	CGCGTCCGATCCGACCGCGCGGA	9568
QY	338	CGTTTACGTCA	CGCGCCATTCGACCTGCTGC	CGCCGCGGAAATACAGCGCACCGTTCACACG	397
Db	9569	CGGTACCGAAATG	CCCGCTGACCTGACCGTTGCGCGAAGCGCGCGCTTCGCCGATCGTCT		9628
QY	398	TGCTCGATGCTCGCT	CCGACCAAGAGGATTCGACGTTGTGCGGATTAACGCGAGGAA		457
Db	9629	TCTTCACCGCTTACTA	CGGGCTGTGTCGACCTTCGGCGGCGCTCCGCGCGGGAGAAAGTGC		9688
QY	458	TCCGATCGCGCGCATTC	ACGCTCTGTTGAAGTTC	CGGCGCGACGATGTATCCAGAAAGTTC	517
Db	9609	TGATTCACGAGACCGAC	CGCGCGGTGTCGACATAGCCCGCGATTCACGCTGCGCGCACCTCG		9748

Oy	518	GTGCCTTCGGCTCGGCGACTGCGCGGCGCTCGGCGCTGGATTGGTGCCCGCCAGTGCAGT	577
Db	9749	GCCGCCAGGTGTTTCGCCACGCGGCGAGCCCGGCGAATGGGGAAGTCTGCGGGGCTTCGTT	9808
Oy	578	AGGAGACCAAGACCTCGTGTGCTCGCTCGTCAACCGAGGGGCTCGCGCTGCTCATGACGTCC	637
Db	9809	TCGACGACGAGCAATCGCTCTCTCCGCGACGCTGGAATTGG--AGAACCGATTCCGCC	9865
Oy	638	TCGATTGACGGCGCGAGAACCCGCTCGAATAATGACGCTTTGACGATGCTGCTTCAGGCCG	697
Db	9866	GGAATGACGTCGTGCTGAGACTCGCTCGCGCAAGAGATTGTCGACGCGTGTGCTGCGGCTGC	9925
Oy	698	AGGCCGACGCGACACAGGCTGAGACGAGAGGCTGTCGCGCTCGTGCGGTGGGATTATCG	757
Db	9926	TGGCGCGAGGCGCGCGCGTTCGTGGAAATGGCGAAGCCACATTCCTGACGCGGAGCAGG	9985
Oy	758	CTGCTGACACCGAATPACACGATCTACCTTATCGCGTTCGCTGAGTCAACTGTCGCGGT	817
Db	9986	TCGGGCGCGCGCAATCCCGGCGTCACTTACCGCGCGTTTCGACTGCTCGACGCCGCGAGCGCC	10045
Oy	818	CG 819	
Db	10046	CG 10047	

```

RESULT 9
US-60-501-821-24
; Sequence 24: Application US/60501821
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: McAlpine, James
; TITLE OF INVENTION: POLYENE POLYKETIDES AND METHODS OF PRODUCTION
; FILE REFERENCE: 3010-1P
; CURRENT APPLICATION NUMBER: US/60/501,821
; CURRENT FILING DATE: 2003-09-11
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 73599
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
US-60-501-821-24

```

```

Query Match          4.3%; Score 54; DB 7; Length 73599;
Best Local Similarity 45.0%; Pred. No. 0.00038;
Matches 244; Conservative 0; Mismatches 295; Indels 3; Gaps 1;

Dh      17912 ACCGGGCTTCGCGCTCTTTCGACGGGGCGGTCGACCGGTTCGGACCGGGCGGA 17971
Qy      278  ACGGATGTTCCGGGAGTCCCGCGGAGATATCACCTCGGGTCGCGACGTCGTCAACCCGT 337

Dh      17972 CCGTCACGGAATAGCCCGGTGACTGACGTCGCGACGCGGCGCTTCGCGGTCTCT 18031
Qy      338  CGTTTACGTACCGGGCCATCGACTCTCTGCGCGCGCAAAATACAGCGCACCGTCGACACG 397

Dh      17972 CCGTCACGGAATAGCCCGGTGACTGACGTCGCGCGGGCTTCGCGCGGGAGAAAGTGC 18031
Qy      398  TGCCTGATGCTCGCTCCGGAACAAGAGAGTTGCACTGTGTGGGGATTACGGGAGGAA 457

Dh      18032 TCCTCACCGGCTACTAGGGGCTGTGCACTTCGCGGGGCTTCGCGCGGGAGAAAGTGC 18091
Qy      458  TCCCGATGCCGGGATCAGCGCTCTGTGAAGGTTCCGGGCGGAGTGTGACGAAAGTTCC 517

Dh      18092 TGATTCACGAGAGCAACCGGCGGTGTGGGCAATGGCCCGGTTCACGCTGGCCCGCACCTTG 18151
Qy      518  GTGCTTCGGGCTCGGCACTGCGCGCGCGCGCTCGGCTGGGATTGTGATGCCCAAGTTCGATG 577

Dh      18152 GCGCCAGAGGTTCGCCACCGGCAAGCCCGGCAATGGGAATGCTGGGGGGGCTCGGTT 18211
Qy      578  AGAGAGACCAAGACCTTGTCGTCGTCCTGTCAACCGAGGGGCTTCGCGCTTCATGACGTCC 637

Dh      18212 TCGACGACGAGCAATGCGCTCTCCCGGACGCTGACATTTCG---AGGACCGGTTTCGGCC 18268
Qy      638  TCGATGAGCGGCGGACGAAACCGGCTCGAAATATACGTTTGACGATGCTGCTTCAGCGCG 697

```

```
Db 18269 GGATGAGCTGCTCTGAGCTCTGCGCCAGAGCTTCGACGCGTCTGCGCGCTGC 18328
Qy 698 AGCCGACGCGAGCAGAGCTGAGCAGAAAGAGCTGCGCGCTCGTGGAGTATATCG 757
Db 18329 TGGGGGAGGGGGCGCGCTTCGAGAGATGGGCAAGCAGACTCCGTGACGGACGAGG 18388
Qy 758 CTGCTGGGACCGATACAGATCTTATGCTGCTGCTGCTCAACCTGCTCGCGT 817
Db 18389 TCGGGCGCGGATCCGCGCGTACCTACCGCGCGTTCGACTGCTCGACCGCGGACGCG 18448
Qy 818 CG 819
Db 18449 CG 18450

RESULT 10
US-10-084-846A-77
; Sequence 77, Application US/10084846A
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TRETZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 77
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avilABCI dna: partial sequence of coding strand 1; nucleotide 1
; US-10-084-846A-77

Query Match 3.9%; Score 49.4; DB 6; Length 972;
Best Local Similarity 46.5%; Pred. No. 0.0017;
Matches 195; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

Qy 288 CGGGCTGCGCGCGAGATCACGCTCGGGTCCGCAAGCTCGTCAACCGTGTTCATGTC 347
Db 504 CGGCTTCGACCGCGCCCAACCGCAGAGGTCTGGAAGATGTCCTGTCGATGTCAGCA 563
Qy 348 AGCGCCATTCGACGCTGCGCGCGGAAATACAGGGCACCGTTCGACGCTGCTGATGC 407
Db 564 GGGGACCAACCGCTGCTACACGACGACTGAGAGAGCGCGACCGGCTGCGGACCA 623
Qy 408 TCGCTCCGAGACAAGAGATTGACGTTGCGGAGATTACGCGAGAGGATCCGATGCG 467
Db 624 GATCTCGGTATTCAGACGCGCGCGCGGCTGCTGAGGGACACCCCGAGACCTTAAATGC 683
Qy 468 CGCGATACGCGCTGTGTGAAGTTCCGCGCGAGTGTGACGAAATTCGTCGCTCGG 527
Db 684 CTTGCTGGGCGCGACCGGATGCTGTGACCTGCGCGGACCGGAGAC---CTGTGCGC 740
Qy 528 CTCGCGGACTGCGCGCGCGCTGCGGCTTGTGTCGCCGAGTTCGATGAGAGACCA 587
Db 741 CGCGCGGCTAGGTGCGCGCTTACCGGAGACCGAGCGCGGAGTGAAGCGCGAGCGCG 800
Qy 588 GACCTGTGCTGCTGCTGCTGACGAGGAGGCTCGCGCTGCTCATGACGCTCTGATGAGCG 647
Db 801 CGGTATTCAGCGCGCGGTATTCAGGACCGGCTCGGACCTCGTGAAGATGTCAGCTGCT 860
Qy 648 GCGCAGAAACCGCTTCGAAATGACGTTTACGATGCTGCTTCAGCGCGGAGCGGACG 706
Db 861 GCGCGGAGAGGCGACCGGCTGTGTGATGTGTGCTGCGCGCGCGGACGCTGAGTAGCG 919
```

```
RESULT 11
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TRETZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; US-10-084-846A-1

Query Match 3.9%; Score 49.4; DB 6; Length 59816;
Best Local Similarity 46.5%; Pred. No. 0.0063;
Matches 195; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

Qy 288 CGGGCTGCGCGCGAGATCACGCTCGGGTCCGCAAGCTCTGCAACCGTGTTCATGTC 347
Db 34691 CGGCTTCGACCGCGCCCAACCGCAGAGGTCTGGAAGATGTCCTGTCGATGTCAGCA 34750
Qy 348 AGCGCCATTCGACGCTGCTGCGCGCGGAAATACAGGACCGCTGACGCTGCTGATGC 407
Db 34751 GGGGACCAACCGTGTGCTGCTACACGACGACTGAGAGAGCGCGCTGCGCGGACCA 34810
Qy 408 TCGCTCCGAGACAAGAGATTGACGCTTGTGCGGAGTTACGGAGGGAATCCGATGCG 467
Db 34811 GATCTCGGTATTCAGACGCGGCGCGCGGCTGTCGAGGGACACCCCGAGACCTTAAATGC 34870
Qy 468 CGCGATACGCGCTGTGTGAAGTTCCGCGCGAGTGTGACGAAATTCGTCGCTGCTCGG 527
Db 34871 CTTGCTGGGCGCGGACCGGATGCTGAGACCTTGGCGGACCGAGAC---CTGTGCGC 34927
Qy 528 CTCGCGGACTGCGCGCGCTGCGGCTGCGGCTTGTGTCGCCAGTTCGATGAGAGACCA 587
Db 34928 CGCGCGGCTGAGGTGCGCGCTTACCGGAGACCGGAGTGTGAGCGCGGAGCGCG 34987
Qy 588 GACCTGTGCTGCTGCTGCTGACGAGGAGGCTGCGGCTGCTCATGACGCTCTGATGAGCG 647
Db 34988 CGGTATTCAGCGCGCGGTATTCAGGACCGGCTGCGGACCTGTCGAGATGTCAGCTGCT 35047
Qy 648 GCGCAGAAACCGCTTCGAAATGACGTTTACGATGCTGCTTCAGCGCGGAGCGGACG 706
Db 35048 GCGCGGAGAGGCGACCGGCTGTGTGATGTGTGCTGCGCGCGCGGACGCTGAGTAGCG 35106

RESULT 12
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TRETZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
```

PRIOR APPLICATION NUMBER: DE 101 09 166.4  
PRIOR FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: Patent Ver. 3.2  
SEQ ID NO 2  
LENGTH: 59816  
TYPE: DNA  
ORGANISM: Streptomyces viridochromogenes  
US-10-084-846A-2

Query Match 3.9%; Score 49.4; DB 6; Length 59816;  
Best Local Similarity 46.1%; Pred. No. 0.0063;  
Matches 195; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 288 CGGCGTCCCGCGGAGATGACGCTGGGTCGCAAGCTGTCACCCGCTTTACGTC 347  
DB 25126 CGGCGTCCCGCGGAGATGACGCTGGGTCGCAAGCTGTCACCCGCTTTACGTC 25067  
QY 348 AGCGCGCATGACCTGCTGGCGCGGAAATACAGCGCACCGTGCACGAGTGTGATGC 407  
DB 25066 GGGCACCACTGCTGCTGCTACCAAGCACTGAGAGAGCGGACCGCTGCGCCAGCA 25007  
QY 408 TCGCTCCGACAGAGAGATGACGCTGTCGCGGATTAAGCGAGAGAAATCCGATGCG 467  
DB 25006 GATCTGATGATGACGCGGCGCGGCGGATGATGAGGAGACCCCGAGACCTGAAATC 24947  
QY 468 CGGATCAGCGCTCTGTTGAAGTTCCGCGCGAGTGTGACGAGAAATTCGTCGCTCGG 527  
DB 24946 CCGTGGCGCGCGGAGCGGATGCTGTCGACCTGCGCGCGCACGAGAC---CTGTGCGC 24890  
QY 528 CTGCGGACATGCGCGCGCGCTGCGGATTTGATGCCCGAGATGAGAGACAA 587  
DB 24889 CGCGCGCTCAGTGTGCTCCGCTTTCAGGAGACCGAGCGGATGAGCGCGAGCGCG 24830  
QY 588 GACCTGTGCTGCGCTGCTGACCGAGGCGCTGCGCTGCTCAATGACGCTCTGATGAGCG 647  
DB 24829 CGTATCAGCGCGCGGATGAGCGAGCGGATGCGCGACCTGTCGATGATGATGCTGCT 24770  
QY 648 GCGGAGAAACCGCTGAAATGAGCTTTGACGATGCTGCTTCAAGCGCGAGCGCGAGC 706  
DB 24769 GCGCGGAGCGGAGCGGATGATGATGCTGCTGCTGCGCGCGCGCGCGCGCTGATGACG 24711

## RESULT 13

PCT-US03-11231-190  
Sequence 190, Application PC/TUS0311231  
GENERAL INFORMATION:  
APPLICANT: Corixa Corporation  
APPLICANT: Day, Craig H.  
APPLICANT: Hosken, Nancy A.  
APPLICANT: Parsons, Joseph M.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
FILE REFERENCE: 210121.53801PC  
CURRENT APPLICATION NUMBER: PCT/US03/11231  
CURRENT FILING DATE: 2003-04-09  
NUMBER OF SEQ ID NOS: 267  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 190  
LENGTH: 9369  
TYPE: DNA  
ORGANISM: HSV2  
PCT-US03-11231-190

Query Match 3.9%; Score 48.8; DB 1; Length 9369;  
Best Local Similarity 46.1%; Pred. No. 0.005;  
Matches 201; Conservative 0; Mismatches 232; Indels 3; Gaps 1;

QY 313 CGGCTCCGAGAGTGTGCAACCGTGTTCATGTCAGCGGCATGACCTGCTGCGCGCC 372  
DB 4489 CGCGATCGCTGACCGCATCCCGGCGGAGCTGTCACCGGCGCTCGCGCGCGCGCTG 4548  
QY 373 GAAATACAGCGACCGTGCACGATGCTGCTGCTGCGCGAGAGAGTTGAC 432

DB 4549 GACCTGCGCGCGCGCGCGGAGATATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4608  
QY 433 GTTGTCCGAGATTAAGCGAGAGAAATCCGATGCGCGGATGACGCTGCTGTTGAAGTT 492  
DB 4609 GAGCTGTGTCGAGACCGGAGGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4668  
QY 433 CGGCGGATGTCAGAGAAATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552  
DB 4669 GCGGCGGCTTGAAGACCGGTCAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4728  
QY 553 GTGCGTTGTCG 609  
DB 4729 GTGCGCTTGTGATCCCGGAGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4788  
QY 610 GAGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669  
DB 4789 CAGCAGCG 4848  
QY 670 GACCTGTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729  
DB 4849 CGCGAGCG 4908  
QY 730 CTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 745  
DB 4909 GTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4924

## RESULT 14

PCT-US03-11231-247  
Sequence 247, Application PC/TUS0311231  
GENERAL INFORMATION:  
APPLICANT: Corixa Corporation  
APPLICANT: Day, Craig H.  
APPLICANT: Hosken, Nancy A.  
APPLICANT: Parsons, Joseph M.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
FILE REFERENCE: 210121.53801PC  
CURRENT APPLICATION NUMBER: PCT/US03/11231  
CURRENT FILING DATE: 2003-04-09  
NUMBER OF SEQ ID NOS: 267  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 247  
LENGTH: 9369  
TYPE: DNA  
ORGANISM: Herpes simplex virus  
PCT-US03-11231-247

Query Match 3.9%; Score 48.8; DB 1; Length 9369;  
Best Local Similarity 46.1%; Pred. No. 0.005;  
Matches 201; Conservative 0; Mismatches 232; Indels 3; Gaps 1;

QY 313 CGGCTCCGAGAGTGTGCAACCGTGTTCATGTCAGCGGCATGACCTGCTGCGCGCC 372  
DB 4489 CGCGATCGCTGACCGCATCCCGGCGGAGCTGTCACCGGCGCTCGCGCGCGCGCTG 4548  
QY 373 GAAATACAGCGACCGTGTGCAACCGTGTTCATGTCAGCGGCATGACCTGCTGCGCGCC 432  
DB 4549 GACCTGCGCGCGCGCGCGGAGATATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4608  
QY 433 GTTGTCCGAGATTAAGCGAGAGAAATCCGATGCGCGGATGACGCTGCTGTTGAAGTT 492  
DB 4609 GAGCTGTGTCGAGACCGGAGGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4668  
QY 433 CGGCGGATGTCAGAGAAATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552  
DB 4669 GCGGCGGCTTGAAGACCGGTCAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4728  
QY 553 GTGCGTTGTCG 609  
DB 4729 GTGCGCTTGTGATCCCGGAGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4788

Oy	610	GAGGGGCGTCGGCGTCGTCATGACGTCTCGANTAGCGCGCAGGAAACCGGCTGGAAAT	669
Db	4189	CAGAGCGCGCGGAGGCCACCGAGCGGGTCAATGGCGGGCTTGGCGGAGGCGCTGTGGCGCG	4848
Oy	670	GACGCTCTTGACGATGCTGCTCTTCAGGCCGAGGCCGACGCGACAGGCTGANGACGAAAGAG	729
Db	4849	CGCGAGCGCCGGGGCCAGATCGAGGCCGAGGGCTTGGCCAACTTCAGACCATGCTGAAG	4908
Oy	730	CTGGTCCGCGCTCGTGG	745
Db	4909	GTGGTCCGCCGATCCCGG	4924

**RESULT 15**

```

US-10-425-114A-14511
? Sequence 14511, Application US/10425114A
? GENERAL INFORMATION:
? APPLICANT: Liu, Jingdong
? APPLICANT: Zhou, Yihua
? APPLICANT: Kovalic, David K.
? APPLICANT: Screen, Steven E.
? APPLICANT: Tabaska, Jack B
? APPLICANT: Cao, Yongwei
? TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
? TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
? FILE REFERENCE: 38-21(5313)B
? CURRENT APPLICATION NUMBER: US/10/425,114A
? CURRENT FILING DATE: 2003-04-28
? NUMBER OF SEQ ID NOS: 73128
? SEQ ID NO 14511
? LENGTH: 1036
? TYPE: DNA
? ORGANISM: Chlorella vulgaris
? FEATURE:
? OTHER INFORMATION: Clone ID: LIB191-016-B10_FLI
US-10-425-114A-14511

```

Query Match	3.8%;	Score 48;	DB 6;	Length 1036;
Best Local Similarity	48.9%;	Pred. No. 0.0041;		
Matches 129;	Conservative	0;	Mismatches 135;	Indels 0;
			Gaps	0

QY	448	GCGAGGGGAATCCCGATGCGGCAGTCAAGCGCTCTGTGAAGTTGCCGCGAGTGAC	507
Db	116	GGGAGGAGATTTCATCTTTCACACTTCCTCGGCCAAGGCCATTGAAGGCGCTGAAC	175
QY	508	GAGAAGTTCCGTGCTCTTCGGCTCGCGGACTGCGCGCGCTCGGCGTGGGTTCGCCCC	567
Db	176	GGCACCGGCGCGCACACCACCACTACTCACAAGAGAGCATCGCAAGACGGCGGCATGGCT	235
QY	568	CAGGTCGATGAGGAGAACCAAGACCCTGGTTCGGTCCGTTCAACGAGGGGCTTCGGCTGCTC	627
Db	236	GCGATTGAGAGCGCATTCAGAAAGCTGTCCAAATGTCACCCGGAGACATTGGCGCACTAC	295
QY	628	CATGACGTCCTCGATGAGCGGCGCAGAAACCGGCTTCGAAATAATGACCTTTGACATGCTG	687
Db	296	GGGTACGGCAACGAGGAGCGGTGTAACCGGCAAGCAAGACATGGAATCAACACCTTC	355
QY	688	CTTAGGCGCCGAGCCGACGGCAGC	711
Db	356	CGTACGGGCTGGCGACCGCGAGC	379

Search completed: October 4, 2003, 12:01:00  
Job time : 233.454 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2003, 03:31:13 ; Search time 289.856 Seconds  
(without alignments)  
11088.586 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_62369\_63628

Perfect score: 1260  
Sequence: 1 atgcacacagagcgaacgaa.....agccctccaaagctcgatag 1260

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	100.0	68750	13	US-10-014-717-1
2	151.4	12.0	1227	12	US-10-214-446-3
3	149.4	11.9	1233	14	US-10-205-032-7
4	149.4	11.9	60196	14	US-10-205-032-1
5	148.8	11.8	1164	14	US-10-156-761-6226
6	148.8	11.8	9025608	14	US-10-156-761-1
7	146.4	11.6	1197	14	US-10-156-761-409
8	146.4	11.6	100000	14	US-10-156-761-15103
9	146.4	11.6	9025608	14	US-10-156-761-1
10	142.6	11.3	1215	12	US-10-214-446-5
11	135	10.7	1248	12	US-10-214-446-55
12	134.8	10.7	1248	12	US-10-214-446-5
13	132	10.5	1197	14	US-10-156-761-2364
14	131.2	10.4	1179	12	US-10-214-446-31
15	129.4	10.3	1278	12	US-10-214-446-19
16	127	10.1	1179	14	US-10-156-761-1975

17	123.4	9.8	1248	14	US-10-156-761-7447	Sequence 7447, Ap
18	123	9.8	1251	10	US-09-861-289-38	Sequence 38, Ap1
19	123	9.8	1251	10	US-09-860-846-38	Sequence 38, Ap1
20	123	9.8	1251	11	US-09-988-384B-38	Sequence 38, Ap1
21	123	9.8	1251	11	US-09-836-821-38	Sequence 38, Ap1
22	123	9.8	5970	11	US-09-793-708-21	Sequence 21, Ap1
23	123	9.8	5970	12	US-10-201-365-11	Sequence 11, Ap1
24	123	9.8	5970	12	US-10-160-539-21	Sequence 21, Ap1
25	120.4	9.6	1191	14	US-10-156-761-7165	Sequence 7165, Ap
26	120	9.5	1215	12	US-10-214-446-39	Sequence 39, Ap1
27	119.8	9.5	1209	14	US-10-205-032-9	Sequence 9, Ap1
28	119.8	9.5	60196	14	US-10-205-032-1	Sequence 1, Ap1
29	119.6	9.5	1215	14	US-10-156-761-4523	Sequence 4523, Ap
30	118	9.4	1212	14	US-10-156-761-2881	Sequence 2881, Ap
31	118	9.4	125746	14	US-10-156-761-15102	Sequence 15102, A
32	116	9.2	1221	12	US-10-214-446-1	Sequence 1, Ap1
33	115	9.1	987	12	US-10-132-134-24	Sequence 24, Ap1
34	115	9.1	1248	14	US-10-156-761-408	Sequence 408, Ap1
35	115	9.1	1700	12	US-10-132-134-22	Sequence 22, Ap1
36	114.6	9.1	1428	12	US-09-953-348-124	Sequence 24, Ap1
37	114.4	9.1	1233	12	US-09-267-255-24	Sequence 24, Ap1
38	114.4	9.1	1233	14	US-09-553-348-74	Sequence 74, Ap1
39	114.4	9.1	12249	14	US-10-267-255-74	Sequence 74, Ap1
40	114.4	9.1	18331	12	US-09-953-348-96	Sequence 96, Ap1
41	114.4	9.1	18331	14	US-10-267-255-96	Sequence 96, Ap1
42	114.4	9.1	1293	14	US-10-145-415-7	Sequence 7, Ap1
43	111.8	8.9	1293	14	US-10-145-415-19	Sequence 19, Ap1
44	111.8	8.9	1293	14	US-10-145-415-19	Sequence 19, Ap1
45	111.4	8.8	1428	14	US-10-145-415-21	Sequence 21, Ap1

## ALIGNMENTS

RESULT 1  
US-10-014-717-1  
; Sequence 1, Application US/10014717  
; Publication No. US20020192778A1  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Finkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF BROTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/10/014,717  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US/09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-10-014-717-1  
Query Match 100.0%; Score 1260; DB 13; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCACAGAGCAGCAAGCAATCAAGAGACCAAGAGCTGTTTCGACTTCAACCCGTC 60  
DB 62369 ATGCACAGAGCAGCAAGCAAGCAATCAAGAGACCAAGAGCTGTTTCGACTTCAACCCGTC 62428  
QY 61 GCGCTGGAGTACGCGAGACCCGTTCCCGCGATGACGCGCTGAGAGAGCAACCC 120  
DB 62429 GCGCTGGAGTACGCGAGACCCGTTCCCGCGATGACGCGCTGAGAGAGCAACCC 62488  
QY 121 ATCTTACTGAGTGAAGGCGCTGCTGCTCTCAACCCGATNACGAGCTGCGCG 180

```
Db 62489 ATCTTCTACTGGAGATGAAGGCGGCTCTGAGGTCCTACCCGATACCGACGATGTCGCG 62548
Qy 181 GTGTTCCGGAGCAAGCGTTTCGCGGTCACTGAGAGAACTGGGAATCGAGCGCGAGTAC 240
Db 62549 GTGTTCCGGAGCAAGCGTTTCGCGGTCACTGAGAGAACTGGGAATCGAGCGCGAGTAC 62608
Qy 241 TGTGCGGCAATCCCGAGCTCAGGCAATGAGAAAGTATGAGATGTTGGGCTGCGCGCG 300
Db 62609 TGTGCGGCAATCCCGAGCTCAGGCAATGAGAAAGTATGAGATGTTGGGCTGCGCGCG 62668
Qy 301 GAGGATCAGCGTCCGGTCCGCAAGCTCTCAACCCCTGCTTACGTCAAGCGCCATCGAC 360
Db 62669 GAGGATCAGCGTCCGGTCCGCAAGCTCTCAACCCCTGCTTACGTCAAGCGCCATCGAC 62728
Qy 361 CTGCTCGCGCGGAAATACAGCGCACCGTGCACACTCTCTGATGCTCTGCTCCGAGCAA 420
Db 62729 CTGCTCGCGCGGAAATACAGCGCACCGTGCACACTCTCTGATGCTCTGCTCCGAGCAA 62788
Qy 421 GAGGATGAGAGTTGTTGGGGAATTAAGGAGGAGGAATCCGATGCGCGCGATCAGCGCT 480
Db 62789 GAGGATGAGAGTTGTTGGGGAATTAAGGAGGAGGAATCCGATGCGCGCGATCAGCGCT 62848
Qy 481 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGTGCGTCCGCTCGCGACCTGCG 540
Db 62849 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGTGCGTCCGCTCGCGACCTGCG 62908
Qy 541 CCGCGCTCGCGCGTGGGTTTGTGTCGCCAGTGTGATGAGAGAACCAAGACCTGTGTCGCG 600
Db 62909 CCGCGCTCGCGCGTGGGTTTGTGTCGCCAGTGTGATGAGAGAACCAAGACCTGTGTCGCG 62968
Qy 601 TCCGTACAGAGGAGGCTCGCGCTGCTCCACATGACGTCTGTGATGAGAGGAGGAGGAGGAG 660
Db 62969 TCCGTACAGAGGAGGCTCGCGCTGCTCCACATGACGTCTGTGATGAGAGGAGGAGGAGGAG 63028
Qy 661 CTGCAAAATGAGCTCTTGAAGATGCTGCTTCAAGCGCGAGCGAGCGAGCGCTGAC 720
Db 63029 CTGCAAAATGAGCTCTTGAAGATGCTGCTTCAAGCGCGAGCGAGCGAGCGCTGAC 63088
Qy 721 ACGAAGAGCTGTGTCGCGCTCGTGGGTGCGATTAATGCTGCTGCTGCAACCGATACACATG 780
Db 63089 ACGAAGAGCTGTGTCGCGCTCGTGGGTGCGATTAATGCTGCTGCTGCAACCGATACACATG 63148
Qy 781 TACCTTATGCGGTTGCTGAGTCAACCTGCTGCGGTCGCGAGGAGGCTCGAGAGTGTG 840
Db 63149 TACCTTATGCGGTTGCTGAGTCAACCTGCTGCGGTCGCGAGGAGGCTCGAGAGTGTG 63208
Qy 841 AAGGCGGAGCGCGGCTCATAGAGAAAGCGGCTCGATGAGGTGCTCGCTTCAATATC 900
Db 63209 AAGGCGGAGCGCGGCTCATAGAGAAAGCGGCTCGATGAGGTGCTCGCTTCAATATC 63268
Qy 901 CTCAGAAATGGAATCTGTGCTTTCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
Db 63269 CTCAGAAATGGAATCTGTGCTTTCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 63328
Qy 961 AAGAAAGGAGGAGTGTCTTTCCTGATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Db 63329 AAGAAAGGAGGAGTGTCTTTCCTGATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 63388
Qy 1021 TCCAGGCGCAGAGCTGTTTATGATGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 63389 TCCAGGCGCAGAGCTGTTTATGATGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 63448
Qy 1081 GAGCCCATGTCGCGCGGAGTGTCTTTCGCTGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
Db 63449 GAGCCCATGTCGCGCGGAGTGTCTTTCGCTGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 63508
Qy 1141 ACCATCTCCGATGTTTCCCGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 63509 ACCATCTCCGATGTTTCCCGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 63568
Qy 1201 CCGCGCTTCGAGAACATCGAATCACTCAAGCTATCTTGAAGGCTCTCAAGAGCTGATAG 1260
Db 63569 CCGCGCTTCGAGAACATCGAATCACTCAAGCTATCTTGAAGGCTCTCAAGAGCTGATAG 63628
```

```
RESULT 2
US-10-214-446-3
; Sequence 3, Application US/10214446
; Publication No. US20030180742A1
GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Bacterial
US-10-214-446-3
Query: Match 12.0%; Score 151.4; DB 12; Length 1227;
Best Local Similarity 49.9%; Pred. No. 5.7e-34;
Matches 513; Conservative 0; Mismatches 486; Indels 30; Gaps 4;
Qy 147 CTGAGTCTCAACCGGATACACAGACGTGTGCGGCTGTTCGCCGACGAACGCTTGCGCGT 206
Db 138 CTGATGTGTACCGGATACGACGAGCGGTGCGCGCTCAACCGACCCCGCTCAGAG 197
Qy 207 CAGTGAGAGAGTGGAGATGAGCGCGGAGATGCTGCGGCAATTCGCGAGCTCAGCGA 266
Db 198 CAGCGCGGTGCGGATACGAGCTGAGAGAGA--GATGCGCACAGAGGAGGACCA 254
Qy 267 TATGAGAGAGTACGATGTTGCGGCTGCGCGGAGGATCAAGCTCGGCTCGGAGCT 326
Db 255 CGTCTGATGACACATGCTGTGCGCACGCGGAGGACCAACCGGCTGCGCACT 314
Qy 327 CGTCAACCGTGTGTTAGTACGCGGCATGACCTGTGCGCGCGGAAATACAGCCAC 386
Db 315 CCTTCGAAGGCTTACCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 374
Qy 387 CGTCAACGCTGCTGATGCTGCTGCGGACCAAGAGGATTCAGATGTCGGGATTA 446
Db 375 CACCGACGCTTCTGACGCGGTCGCGCGCGCGGATTCGCGGACCTGCTGCGAGTT 434
Qy 447 CCGGAGGAGATCCGATGCGCGGATGACGCTGCTGTTGAAGTTCCGCGCGAGTGA 506
Db 435 CGCGTGCAGCTCCCATAGGCGGTACTCAGGACCTCATCGGACCGCGCGAGGAGCA 494
Qy 507 CGAAGATTCGCTGCTGCTGCGGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566
Db 495 GCCCGATGCGCGC-----CTGCGGTCGCGCTCATCATCCGCG 536
Qy 567 CAGGTCATAGGAGACCAAGACCTGTGCGCTGCTGCTCAACGAGGAGGCTGCGCTGCT 626
Db 537 GACACACCCCGAGGAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 596
Qy 627 CCATGACGCTCATGATGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 686
Db 597 CGAGCGTTGATCGCGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 650
Qy 687 GCTTACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 746
Db 651 GCTTACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 710
Qy 747 TCGATTAATGCTGCTGCGGACGATACGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
```

Db 711 CTTGCTCAGCTGCGCGGCAACGACGACCGCCGCTGATCGCAACGCGGCTGACAGC 770  
Qy 807 CTTGCTGAGTGGCGCGGCGCTGAGTGGAGCGCGGCGCTGATGAGAA 866  
Db 771 CTTGCTGAGTGGCGCGGCGCTGAGTGGAGCGCGGCGCTGATGAGAA 830  
Qy 867 CCGCGTGCATGAGTGGTCTCGCTGAGCAATATCTTCAAGAAATAGAACTGTGCTTTCG 926  
Db 831 CCGCATGAGGAACTCTGCGCTGACGAGGCGCGGCTGAGCGGCGCTGCGCGCTTAC 890  
Qy 927 CAGGAGGACCTGAGTACTGCGGCGCATTCATCAAGAAAGGAGATGTCTTCTCT 986  
Db 891 CACGACCCGTAAGATGCGGCGGCTGACCGTACCGCGCGGCGAGATGATCATGCG 950  
Qy 987 GATCCGAGCGCGCTGAGATGAGATGAGTATTCCTCCAGGCGAGCGTGTGA---TGT 1043  
Db 951 GCTGCGCGCGCAATCGGACCGCGCGCTGACGACGCTCCGACATCTGAGCTTGC 1010  
Qy 1044 GCGACGAGCAACGCGCGGCGAGCTGCGTACGCTGAGAGCGCGCGCTGCTGCGGAGT 1103  
Db 1011 ACGCGTGAAGTGGCGCAACGCTGCTTGGCGCATGGCGTGCATCTTCTGCTGGTGC 1070  
Qy 1104 GTCCCTTCTGCTGCTGAGGCGGAGATGCGGTGGGCAACATCTTCCGATGCTCCGA 1163  
Db 1071 GCCGCTGCGCGCGCGGAGCGCGGATCGCATCGGACCGCTGCTGCGCGCTTCCCGA 1130  
Qy 1164 GATGAGCT 1172  
Db 1131 TCTGCGCT 1139

RESULT 3  
US-10-205-032-7  
; Sequence 7, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Stafia, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-205  
; CURRENT APPLICATION NUMBER: US/10/205, 032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 1233  
; TYPE: DNA  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-7

Query Match 11.9%; Score 149.4; DB 14; Length 1233;  
Best Local Similarity 49.7%; Pred. No. 2.2e-33;  
Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;  
Qy 288 CGGGGTGCGCGCGGAGATCAAGCTCGGCTCGCAAGCTCGCAACCGCTGTTAGCTC 347  
Db 261 CGGATCGACCGCGCGCGGAGCGCGCGCTGCGAGCTGAGAGCGGCTTACCGCC 320  
Qy 348 ACGGCGCATCGACCTGTGCGCGCGCGCAATACAGCGACCGTGCAGCTGCTGATGC 407  
Db 321 CCGCGGATCGCGCGCGAGTGGAGTGGCGAGTACCGCGCGCGAGTGGTGCAGCA 380  
Qy 408 TCGCTCGGACAAAGAGTTCGACGTTGCGGAGTATACCGGAGGGAATCCGATGCG 467  
Db 381 GGTACGACACGAGACCGCATGACATGCGACGACCTCGGTACCGCGCTGCGGTAC 440  
Qy 468 CGCGATCGCGCTGTTGAAGTTCCGCGGAGTGTACAGAGAAATTCGCTGCTGCG 527  
Db 441 GGTATCGCGAGCTGCTGCGGATTCACCAAGAGATCAAGAAATTCGCGAGTGGGT 500  
Qy 528 CTCGCGACCTGCGCGCGCTGCGGTGGTGTGCTGCCAGGTGATGAGAGACCA 587

Db 501 GGACATCATCTTCAGCAACGAGGCTGAGATATCCAACTCCCGGACGATTCACCGA 560  
Qy 588 GACCTGTGCGCTGCTGCAACCGAGGGGCTGCGGCTGCTCATGACGCTCGATGACG 647  
Db 561 GACGTGCGCGCGCGCATGAGAGTGTCCGAATTCCTGTAACCGCAATGCGCACAA 620  
Qy 648 GCGAGGAAACCGCTGCAAAATGACGCTTGTACAGATGCTGTTAGAGCGCGAGCG 707  
Db 621 GCGCGCGGACCG---AAGACGACCTGATAGAGGCTCTGTGCGGCGAGAGTGCAG 677  
Qy 708 CAGAGGCTGAGCAAGAGAGTGTGCGCTGCTGAGGTCGATATGCTGCTGCGAC 767  
Db 678 GCGAGAGCTGACCGACGAGAGATGCTACATGCTGCGCTGCTCACCGCGGCGCA 737  
Qy 768 CGATACACAGATCTACCTTATGCGGCTTGTGCTGCTCAACCTGCTGCGGCTCCGAGG 827  
Db 738 CATCTCCAGCGCGACGCTGCTGCAACCTGTCTGCTGCTGAGAGACCGCGAGG 797  
Qy 828 GCTGAGCTGAGAGCGCGGCGGCTGATGAGAAACGCGCTGATGAGTGTGCTCG 887  
Db 798 ACAGCGCGGCTCGCGCGCGAGCGCGAGCTGTGCGCGGCTGATGAGAGAGCTGCG 857  
Qy 888 CTTCGACATATCTCTCAAGATGAGAACTGTGCTTGGCGAGGAGAGCTGAGACTG 947  
Db 858 CT---ACGAGTCCCGTTCAACTGATCTTCCGATCCTGAACGAGGACCGACATCT 914  
Qy 948 CGGGCATCATCAAGAAAGGAGATGTCTTCTCTGATCCCGAGCGCTGAGAGA 1007  
Db 915 CGGCGACCCCATGCGCAAGGCGCAGATGTGATGCTGATGCTTCCGGAACCGGGA 974  
Qy 1008 TGGAGTGTATTTCTCCAGGCGAGAGTGTGATGTCGAGCGGACGCGGCGAGCT 1067  
Db 975 CACGAGGTGTTCAACGAGACCGGACACCTTGCATCGACGCGAGTGAACAGCACT 1034  
Qy 1068 GCGGTACGTAGAGCGCGCGCATGTCTCCCGGAGTGTCTTCTGCTGCTGAGAGCG 1127  
Db 1035 GCGGTGCGCGACGCGCATTCACACTGCTGCGGCGCTTCTGCGCAGGCTGAGGCGAA 1094  
Qy 1128 GATGCGCGTGGGACCATCTTCCGATGCTTCCCGAGAT 1166  
Db 1095 GGTCTTCTCAACGAGCGCTGCGACGATTCACCGAGTT 1133

RESULT 4  
US-10-205-032-1/C  
; Sequence 1, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianshu  
; APPLICANT: Stafia, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-205  
; CURRENT APPLICATION NUMBER: US/10/205, 032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 60196  
; TYPE: DNA  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-1

Query Match 11.9%; Score 149.4; DB 14; Length 60196;  
Best Local Similarity 49.7%; Pred. No. 4.3e-33;  
Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;  
Qy 288 CGGGCTCGCGCGGAGATCAAGCTCGGCTCGCAAGCTGTCAAACCGCTTATGCTC 347  
Db 3601 CGGATGACCGCGCGCGGAGCGCGCTGCGAGCTGTGAGCCAGAGGCTTACCGCC 3542  
Qy 348 ACGGCGATCGACCTGCTGCGCGCGCAATACAGCGACCGTGCAGCAGCTGCTGATGC 407

```

Db 3541 CCGCCGATCGCCAGATGAGCTGGGATCGGGGAGATACCGCCGAGCTCTGACCA 3482
Qy 408 TCGCTCCGACAAAGAGAGTTCAGCTTTCGGGATTAACGGGAGGAATCCGATGCG 467
Db 3481 GGTACGGACCAAGACCGGATGACATGCGACGACCTTGGGTACCCGCTCGGTGAC 3422
Qy 468 CGCGATCAAGCTCTGTTGAAGTTTCCGCGAGTGTAGAGAAATTCTTCGCTTGG 527
Db 3421 GGTCAATCGCGAGCTGCTCGGCAATTCACCAAGATCAAGAAATTCCGGAGTGGT 3362
Qy 528 CTCGGAGATGCGCGCGCGCTGCGCGGCTTGGGTTTGGGCCCCAGGTGATAGAGACCA 587
Db 3361 GGACATCATCTCAGACAGAAAGGCTGAGTATCCCAACTTCCGAGACATTCACCGA 3302
Qy 588 GACCTGTGCGGTGCGTCAACGAGGAGCTGCGCTGCTCATGACGTCTCGATGAGCG 647
Db 3301 GACGGTGGGCCCCCGCATGAGAGAGTGTCCGAATTCTGTACGCCAGATCGCCACAA 3242
Qy 648 GCGCAGAAACCGGCTCGAAAATGAGCTTTGACGATGCTGCTTCAGGCGGAGCGGCG 707
Db 3241 GCGCCCGCAACG---MAGACGACCTGATACGCGGCTCTGTGCGCGAGAGTGCAGCG 3185
Qy 708 CAGCAGCTGAGACAGAAAGAGTGTGCGGCTGCGGAGTGGGATTAATGCTGTGCGAC 767
Db 3184 GCGCAAGCTGACGACGAGAAAGTGTCAACATGCTGCGCTGCTCAGCGCGGCGA 3125
Qy 768 CGATACCAAGATCTACTTATGCGCTGCTGCTCAACTGTGCGGTGCGCGCGAGCG 827
Db 3124 CATCTTCAGCGCGACGCTGCTCAGCAACTGTTCCTGTGTGAGAGACACCGGACGC 3065
Qy 828 GCTCAGCTGTGGAAGCGGAGCGGCTCATGAGAAACGGCTTGAATGAGTGTGCG 887
Db 3064 AAGGCGCGGTGCGCGCGACCGACGCTGCGCGGCGGTGATGAGAAACGCTGCG 3005
Qy 888 CTTCGACATATCTCAGATAGAACTGTGCTTTCGCGAGAGACCTGAGATGCTG 947
Db 3004 CT---ACGGGTCCCGTTTAACTGCACTTCCGATCTCGAAGAGAGACACGACATCT 2948
Qy 948 CGGGGATCGATCAAGAAAGGAGATGTCTTCTCTGATCCGAGCGCCTGAGAGA 1007
Db 2947 CGGCAACCCCATGCGCAAGGCGCAATGTGTGATGCTGATGCTCGCGCAACCGGCA 2888
Qy 1008 TGGGACTGATTTCTCAGGCGCAGAGTGTGATGTGCGACGAGACAGGCGGCGACT 1067
Db 2887 CACCGAGGTTCACGAGACCGGACACTTCGACATCGACGAGTCAACAAGACCT 2828
Qy 1068 CGCGTACGAGTAGAGCGCCCATGTGCGCGGAGTGTCTTGTGCTGCGCTGAGGCGGA 1127
Db 2827 GGCCTTCCGCGACAGGATCAACACTGCTTGGGCGGCTTCTGGCAGGCTGAGGCGAA 2768
Qy 1128 GATCGCGGTGGGACCACTTTCGTAAGTTTCCCGAGAT 1166
Db 2767 GGTCTTCTCAACAGACGCTCGACCACTTCAACCGATT 2729

```

RESULT 5  
US-10-156-761-6226  
; Sequence 6226, Application US/10156761  
; Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089

```

; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6226
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1164)
US-10-156-761-6226

Query Match      11.8%; Score 148.8; DB 14; Length 1164;
Best Local Similarity 50.0%; Pred. No. 3.2e-33;
Matches 449; Conservative 0; Mismatches 422; Indels 27; Gaps 2;

Qy 284 TGTTCGGGCTGCGCGGAGATCAAGCTCGGGTCCGCAAGCTGTCACCGCTGTTTA 343
Db 223 TGTTCACATGAGCGCGCGGACACACCGCATCCGCGCTGTGTCGGCGGCGCTTCA 292
Qy 344 GGTCAAGCGCATGACCTGTGCGCGCGGAAATACAGCGACCGTGCACGACTGCTCG 403
Db 293 CCTCGTGTGCGTGAACAATCGCGAGCCGTAAGGAGACCGCCACGCGCTTCTCG 352
Qy 404 ATGCTGCTCGGACAAAGAGATTCGACTTGTGCGGAGATTACCGGAGGAAATCCCGA 463
Db 353 ACGACTGCGGAGACGACGCGACACGACTGATGCTGTGACGCGCGCGCTGCGCA 412
Qy 464 TCGCGCGCATGACGCTCTGTTGAAGTTCGCGCGAGATGTGACAGAAATTCCGCTGCT 523
Db 413 TCACGTGATCTGCGACCTGCTCGGCGTCCGAGACAAACCGGCGGAGATTCCGGGCT 472
Qy 524 TCGGCTGCGGACGTCGCGCGCGCTCGGCGTGTGGTTGTGTCGCCAGTTCATAGAGAGA 583
Db 473 GGACGACACCGC-----TCGTACCCCGGACCGCGCGCC 508
Qy 584 CCAAGACCTGTGTCGCTGCTGACCGAGGCGTCCGCTGCTCATAGACGCTCGATG 643
Db 509 CGAGGTGCGCAAGGAAATCGTGTGCTGTGCTGCTTCAACCGGCTCTGCGCG 568
Qy 644 AGCGCGCAGAGAACCGCTCGAAATGACGCTTTCAGCATGCTGCTTCAAGCGCGAGCGG 703
Db 569 ACAAGCGCAAGAACCC---CCGAGACGACGTGCTTCCGACTCATTCGCGGAGAG 625
Qy 704 ACGGACGAGCTGTGACAGAAAGCTGTGCGCTGCTGTGAGTGCATTATGCTGCTG 763
Db 626 AGGGGACCGGCTCACGAGAGACGAGTGAATGCTGCTGCTTCTCATCTCTTTCGCG 685
Qy 764 GCACGATACACGATCTACCTTATCGGCTTGTGCTGCTCAACTGTGCGGCTGCGCG 823
Db 686 GTTACGAGAAACCGTGTGACCTTCAATGAGAAAGCGCTTGTGCTGCGCATTCGG 745
Qy 824 AGGCGCTGAGCTGTGAGAGCGGAGCGGAGCTTATGAGAAACGCGCTGATGAGTGC 883
Db 746 AGCACTGCGCGCGCTCGCGAGAGACCGGACCGGCTGCGCGCGGAGATTG 805
Qy 884 TCCGTTGACATATCTCTCAGATAGAACTGTGCTTGTGCGCAGGACGATCTGAGT 943
Db 806 CCCGCTACGAAAGGCGCGCTGCTGCGCATCCGCGCTTCCGCTGTGGAGACGTGACGA 865
Qy 944 ACTGCGGGGCAATGATCAAGAAAGGAGATGTTCTTCTCTGATTCGCGAGCGCTGGA 1003
Db 866 TCGCGGGGTACACGTGCGCGCGGAGAGACCGTCTGTGCTGTGTCGCGCGCAAC 925
Qy 1004 GAGATGGAAGTGTATTTCTCAGGCGCAGAGCTGTTGATGTGAGACGAGACACGCGGCGA 1063
Db 926 GTGACCGAGCGGCTTCCCGACCCCGACCGGCTGATCTGGGCGGACGCGCGGAGGCC 985
Qy 1064 GCTTCGCTGAGTGAAGCGCCCAATGTGCTGCGCGGAGTGTCTTGTGCTGCTGAGG 1123
Db 986 ACCTGCGCTGCGCGACGCGGCTGCACTGCTGTGGGCGCGCGCTGCGCGCTGAGAGA 1045

```

Qy 1124 CGAGATCGCCGTCGAGCAACATCTTCGTAAGTTCCCGAGATGAAGCTGAAGAAC 1181  
Db 1046 CGAGATCGCCCTCGCGCGCTCTCGAACGTTCCCGAGACTCGCCCTCGCGAAG 1103

RESULT 6  
US-10-156-761-1

; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 11.8%; Score 148.8; DB 14; Length 9025608;  
Best Local Similarity 50.0%; Pred. No. 1.5e-32;  
Matches 449; Conservative 0; Mismatches 422; Indels 27; Gaps 2;

Qy 284 TGTTCGGGCTCCCGCGAGATCAGCTGCGGTCGCGAAGCTGTCACCCGTCGTTA 343  
Db 7508344 TGCTCAACATGAGACGCGCGGACCAACCCGACCTCGGCTCGCGCGGCTTCA 7508403  
Qy 344 CGTACAGCGGCATGACCTGTCGGCGCGCGAATTCAGCGACGTCGACGAGTCGCG 403  
Db 7508404 CCTGCGTCGCGTGAACATCTCCGCGAGCCGTAAGGAGACGCGCCACGCGCTTCG 7508463  
Qy 404 ATGCTCGCTCCGGAACAAGAGATTGACGTTGTCGGGATTACGCGAGGGAATCCCGA 463  
Db 7508464 ACGCACTCGGAGCGAGCGGACCGACCTCATGCTGCTGTAAGCCGCGCTGCGA 7508523  
Qy 464 TGCAGCGATAGCGCTCTGTTGAAGTTCCGCGCGAGTGTGACGAGAAATCCCGA 523  
Db 7508524 TCACCTGATCTGCGACCTGCTCGCGCTGCGGAGCAACCGCGGAGACTTCCGCGCT 7508583  
Qy 524 TCGGCTGCGGAGCTGCGCGCGCTGCGCGCTGCGGTTGTTGTCGCCAGTGTGATGAGA 583  
Db 7508584 GGACCGACCCGCT 7508619  
Qy 584 CCAAGACCTGTCGTCGCTGTCACCGAGGAGTGCCTGCTCCATGACGTCCTGATG 643  
Db 7508620 CGAGCTGCGCAAGGAATCTCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 7508679  
Qy 644 AGCGCGCGAGAACCCGCTGGAATAAGCTGTTGACGATGCTGCTTACGCGCGAGCGG 703  
Db 7508680 ACAAGCGGAAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7508736  
Qy 704 ACGGAGAGGCTGAGCAAGAGAGTGTGCGGCTGTTGAGTTATGCTGCTG 763  
Db 7508737 AGGGGAGCGGCTACCGAGAGCGATGATGCTCTGCTCTCTCTCTCTCTCTCTCT 7508796  
Qy 764 GCACGATACCAAGATCTACTTATGCGCTTCTGCTGCTCAACCTGCTGCGGTCGCGG 823

Db 7508797 GCTACGAGAACACCTGTACCTCATCTGGAACCGCGTAATGCGCTGCTGCGCATCCG 7508856  
Qy 824 AGCGCTCGAGCTGTGAAGCCGAGCCCGGCTCATGAGGAACGCGTGTGATGAGTGC 883  
Db 7508857 AGCAGCTGCGCGCTCTCGCGAGACCGGACCGCTGCGGACCGCTCGGAGATTGC 7508916  
Qy 884 TCCGCTTCGACAAATATCTCTGATAGAACTGTGCTTTTCGACGAGCACTGTAGT 943  
Db 7508917 CCGCTACGAAGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7508976  
Qy 944 ACTGCGGCGCATGCATGAAGAGGAGATGCTTTCTTCTATCCGAGCCCTCGA 1003  
Db 7508977 TCGCGGCGGTACACCTGCTCGCGGAGACCGGTGCTGCTGCTGCTGCTGCTG 7509036  
Qy 1004 GAGATGAGATCTATCTTCCAGGCGACAGCTGTTGATGTCGACGGGACAGCGGCGCGA 1063  
Db 7509037 GTGACCGAGCGGCTTCCGACCCGACCGCTGATCTGCGCGGACCGCGAGGCC 7509096  
Qy 1064 GCTCGCGTACGTAAGAGCCCGCATGTCGCCCCGCGGTTGCTTGTGCTGCTGAGG 1123  
Db 7509097 ACTTCGCGCTGCGGACGAGCGTCACTGCTGCGCGCGCTGCTGCTGCTGCTGAG 7509156  
Qy 1124 CGAGATCGCCGTCGAGCAACATCTTCGTAAGTTCCCGAGATGAAGCTGAAGAAC 1181  
Db 7509157 CGAGATCGCCCTCGCGCGCTCTCGAACGTTCCGAGACTGCTGCTGCGGAAAC 7509214

RESULT 7  
US-10-156-761-409

; Sequence 409, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 409  
; LENGTH: 1197  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1197)  
US-10-156-761-409

Query Match 11.6%; Score 146.4; DB 14; Length 1197;  
Best Local Similarity 50.8%; Pred. No. 1.6e-32;  
Matches 453; Conservative 0; Mismatches 406; Indels 33; Gaps 3;

Qy 279 CGGATTTTCGGGTCGCGCGCGAGATCAACGCTCGGTCGGAAGCTGTCGAACCCGTC 338  
Db 255 CAGCTGTGTTGAGAGAGAGCGGAGACACTTGGACCGCAAGCTGCTGCGCAAGGA 314  
Qy 339 GTTACGTCAGCGCCATTCGACTGTGCGCGCGGAATACAGCGACCGTGCACCACT 398  
Db 315 GTTACCGTACGCGGAGTACGAGGCGCTGCGCGGCAATCAGCGCATTCGTGACGAGCA 374  
Qy 399 GCTGATGCTGCTGCTCGGACAGAGG--AGTTGACGTTGTGCGGAGTTACGCGGAGG 455  
Db 375 CTTGATGATCATCAGAGCGCGGCGCGCTGACCTGTTGAAGACTTTCGCCAAGCG 434

```

Qy 456 AATCCGATCCGCGCATGAGGCTTGTGAAGTTCCGCGGAGTGTGACGAAAGTT 515
Db 435 GGTCGCGTCATGATGATTTCCGACCTGTTGGTGCTCCGTCGACCGCGGAGAGTT 494
Qy 516 CCGTCGCTTCGAGCTCGGAGTACGCGCGCGCTCGGCGTGGGTTTGTGCCCCAGGTGCA 575
Db 495 CAGAGCATTCGCGGAGCGCATGATGCG-----GGTCA 527
Qy 576 TAGAGAGACCAAGACCTGTCGCGCTCGCTCACCGAGGGCTCGCGCTCTCCATGACGT 635
Db 528 CAGAGACGCGCGCGCACCGAGGCGCGCGCATGCGCTCGGCGGAGCTCTCTACACAGCT 587
Qy 636 CCGTCAATGAGCGCGCGAGAAACCCGCTCGAATAATGACGTTCAGCATCTGCTTCAGGC 695
Db 588 TGTTCAGAGGCGCGCGGCGCAACCC--CGGTGACGACGTATCTCGGCGCTGATCAACAC 644
Qy 696 CGAGCGCGAGCGGAGAGGCTGAGCAGAGGAGTGTGCGCGCTCGTGGGTGCGATAT 755
Db 645 CAGAGACCCCGATGATGATGTGTCAGACATGTTTCTCATGAACGCGCGCGGACCTGCT 704
Qy 756 CGCTGTGACCGATACGACGATCTATCGGCTTCGCTGCTCAACCTGCTGCG 815
Db 705 GATAGGCGGCGCACAGACACCGCTGATGATCGGCTCGGTCAGCGCGCTGCTCTGCA 764
Qy 816 GTGCGCGCGAGCGCTGAGCTGTGAAGCCGAGCCCGGCTCATGAGAAACGCGTCA 875
Db 765 CAGCCCGGACGACGTGCGCTGTCGCGAGACCCGTCGTGTGSCAACGCGCTGCA 824
Qy 876 TAGAGTCTCCGCTTCGACAAATCTCTCAGAAATAGAACTGTCGCTTCGCGAGGAGA 935
Db 825 GAGAGCTGTGCGTACTCTACCATCGGCCAGTTCCGCGGCGAGCGGCTGCGCACCGGGA 884
Qy 936 CCGTGAATCTGCGGCGCATTCGATCAAGAAAGGAGATGTTCTTCTGATCCGAG 995
Db 885 CGTGAGCTCGGTGGGTGCGCATCGCCAAAGGCGAGAGTGTGCGCGACGCTCTCGC 944
Qy 996 CGCGCCGAGAGATGGGACTGATCTCCAGGCGCAGACGTTGATGTGACGCGGACAC 1055
Db 945 CGCGCATTCGACCCCGCTTCGTGAGAGACCGGAAAGTTTCAGACATCACTGACGCGC 1004
Qy 1056 GGGCGCGAGCTTCGCTAGCGGTAGAGCGCCCATGTCTCCCGGGGTGTCCCTTGCTCG 1115
Db 1005 CGCGCCGCGATCTGCGCTTGGGCTTGGCGGACACAGTGCATCGGCGACCACTGCGCG 1064
Qy 1116 CCTCGAGCGGAGATCGCGGTGGGACATCTTCCGTAGGTTCCCGAATG 1167
Db 1065 GATCGAGCTGACATGCTCTTCGAGACCTGTTCCGCGCGCTCCGCGTCTG 1116

```

## RESULT 8

```

US-10-156-761-15103
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000

```

```

; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103
Query Match 11.6%; Score 146.4; DB 14; Length 100000;
Best Local Similarity 50.8%; Pred. No. 3.5e-32;
Matches 453; Conservative 0; Mismatches 406; Indels 33; Gaps 3;
Qy 279 CGGATTTGTCGGGCTGCGCGCGGAGATCAGCTCGGTCCGCGAAGCTCTCAACCCGTC 338
Db 80684 CAGCCTGTTGTGGCAGAGACGAGCCGAGACACCTCGGACCGCATGCTGCGCAAGA 80743
Qy 339 GTTACGTACGCGGCATCGACTTGTGCGCGCGGAAATACAGCGCACCGTCAACAGCT 398
Db 80744 GTTCAACGTACGCGGCATCGAGCGCTTGGCGCGGAACTCAAGCGCATCTGTCAGAGCA 80803
Qy 399 GCTCATGCTGCTTCGCGGACAAAGG--AGTTGACGTTGTGCGGATTAACGCGAGAG 455
Db 80804 CCGATGATCATCGAGGCGCGGCGCGCGGCTCGACCTGTGAAGAC

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:46:26 ; Search time 66.7507 seconds  
(without alignments)  
8331.640 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_62369\_63628

Perfect score: 1260

Sequence: 1 atgacacagagcaagcgaa.....agccctccaagctgatag 1260

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/ECTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	100.0	68750	US-09-335-409-1	Sequence 1, Appli
2	1260	100.0	68750	US-09-568-102-1	Sequence 1, Appli
3	1260	100.0	68750	US-09-567-969-1	Sequence 1, Appli
4	1260	100.0	68750	US-09-568-480-1	Sequence 1, Appli
5	1260	100.0	68750	US-09-568-486-1	Sequence 1, Appli
6	1260	100.0	68750	US-09-568-472-1	Sequence 1, Appli
7	1260	100.0	68750	US-09-567-899-1	Sequence 1, Appli
8	1246.2	99.9	71989	US-09-443-501A-2	Sequence 2, Appli
9	145.4	11.5	1314	US-09-252-991A-1265	Sequence 1265, Ap
10	145.4	11.5	1482	US-09-252-991A-1338	Sequence 1338, Ap
11	145.4	11.5	1671	US-09-252-991A-1222	Sequence 1222, Ap
12	123	9.8	1251	US-09-105-537-38	Sequence 38, Appli
13	123	9.8	5970	US-09-320-878-21	Sequence 21, Appli
14	123	9.8	5970	US-09-141-908-11	Sequence 11, Appli
15	123	9.8	5970	US-09-657-440-21	Sequence 21, Appli
16	118.2	9.4	1407	US-09-252-991A-15866	Sequence 15866, A
17	118.2	9.4	1731	US-09-252-991A-15928	Sequence 15928, A
18	118.2	9.4	2502	US-09-252-991A-15767	Sequence 15767, A
19	115.6	9.2	1212	5212296-17	Sequence 17, Appli
20	115	9.1	1221	5212296-16	Sequence 16, Appli
21	115	9.1	1879	5212296-5	Sequence 5, Appli
22	114.4	9.1	1227	US-09-385-028-23	Sequence 23, Appli
23	114.4	9.1	1227	US-09-726-614-23	Sequence 23, Appli
24	114.4	9.1	1233	US-09-266-965-24	Sequence 24, Appli
25	114.4	9.1	12249	US-09-266-965-74	Sequence 74, Appli
26	114.4	9.1	15079	US-09-385-028-1	Sequence 1, Appli
27	114.4	9.1	15079	US-09-726-614-1	Sequence 1, Appli

28	114.4	9.1	18331	US-09-266-965-96	Sequence 96, Appli
29	111.4	8.8	1735	US-08-102-863-10	Sequence 10, Appli
30	111.4	8.8	1735	PCT-US92-10885-10	Sequence 10, Appli
31	109.6	8.7	1224	US-09-266-965-22	Sequence 22, Appli
32	97.4	7.7	50937	US-09-428-517-1	Sequence 1, Appli
33	94.8	7.5	1998	5212296-8	Sequence 8, Appli
34	94.6	7.5	47981	US-09-679-279-1	Sequence 1, Appli
35	86.2	6.8	4403765	US-09-103-840A-2	Sequence 2, Appli
36	86.2	6.8	4411529	US-09-103-840A-1	Sequence 1, Appli
37	82.8	6.6	77536	US-09-410-551B-1	Sequence 1, Appli
38	80.2	6.4	6085	US-09-029-603-4	Sequence 4, Appli
39	77.6	6.2	1188	US-09-266-965-62	Sequence 62, Appli
40	77.6	6.2	53500	US-09-266-965-76	Sequence 76, Appli
41	75.2	6.0	1209	US-09-105-537-21	Sequence 21, Appli
42	75.2	6.0	13613	US-09-105-537-3	Sequence 3, Appli
43	75.2	6.0	38506	US-09-320-878-19	Sequence 19, Appli
44	75.2	6.0	38506	US-09-141-908-1	Sequence 1, Appli
45	75.2	6.0	38506	US-09-657-440-19	Sequence 19, Appli

#### ALIGNMENTS

RESULT 1  
US-09-335-409-1  
; Sequence 1, Application US/09335409  
; Patent No. 6121029  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goetlich, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/335,409  
; CURRENT FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-335-409-1

Query Match	Score	DB 3;	Length	68750;
Best Local Similarity	100.0%;	Pred. No. 9.9e-250;		
Matches 1260; Conservative	0;	Mismatches	0;	Gaps 0;
1	ATGACACAGGAGCAAGCGAATCAGATGAGAGCAAGCGCTTTCGACTTCAAGCGCTTC	60		
62369	ATGACACAGGAGCAAGCGAATCAGATGAGAGCAAGCGCTTTCGACTTCAAGCGCTTC	62428		
61	CGCGCTGGGTACGCGGAGAGCCCGTCCCGCATGAGCGCTTTCGAGAGCAAGCGCTTC	120		
62429	CGCGCTGGGTACGCGGAGAGCCCGTCCCGCATGAGCGCTTTCGAGAGCAAGCGCTTC	62488		
121	ATCTTCTACTGAGATGAAGCGCGCTCTGAGTCTTCTACCCGATACGAGCGTGTGGCG	180		
62489	ATCTTCTACTGAGATGAAGCGCGCTCTGAGTCTTCTACCCGATACGAGCGTGTGGCG	62548		
181	GTGTCGCGGAGCAAGCGCTTTCGCGTCTGAGAGAGTGGGAATCGAGCGCGAGTAC	240		
62549	GTGTCGCGGAGCAAGCGCTTTCGCGTCTGAGAGAGTGGGAATCGAGCGCGAGTAC	62608		
241	TCGTGCGCATTTCCGAGACTGAGCATATGAAGATTCGAGCTTTCGAGCTGCGCGC	300		
62609	TCGTGCGCATTTCCGAGACTGAGCATATGAAGATTCGAGCTTTCGAGCTGCGCGC	62668		
301	GAGATCAGCGTGGGCTCGCAAGCTCGTCAACCGCTGTTTACGTCAGCGGCATCGAC	360		
62669	GAGATCAGCGTGGGCTCGCAAGCTCGTCAACCGCTGTTTACGTCAGCGGCATCGAC	62728		



QY 361 CTGCTGCGCGGAAATACAGCGCACCGTGCACAGCTGCTGATGCTGCTCGGACAA 420  
DB 62729 CTGCTGCGCGCGGAAATACAGCGCACCGTGCACAGCTGCTGATGCTGCTCGGACAA 62788  
QY 421 GAGGAGTTGAGCGTTGTGCGGGAATACGCGGAGGGAATCCGATGCGCGGATCAGCGCT 480  
DB 62789 GAGGAGTTGAGCGTTGTGCGGGAATACGCGGAGGGAATCCGATGCGCGGATCAGCGCT 62848  
QY 481 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGGCACTGCG 540  
DB 62849 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGGCACTGCG 62908  
QY 541 CCGCGCGCTGCGCGGTTTGTGCTGCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGG 600  
DB 62909 CCGCGCGCTGCGCGGTTTGTGCTGCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGG 62968  
QY 601 TCCGTCACGAGGAGGCTGCGCGGCTGCTCAATGAGTGTGACGAGGAGGCGGAGAAACCGG 660  
DB 62969 TCCGTCACGAGGAGGCTGCGCGGCTGCTCAATGAGTGTGACGAGGAGGCGGAGAAACCGG 63028  
QY 661 CTGGAAGATGACGCTTCTTGAAGATGCTGCTGAGGCGGAGGCGGAGGCGGAGGCTGAGC 720  
DB 63029 CTGGAAGATGACGCTTCTTGAAGATGCTGCTGAGGCGGAGGCGGAGGCGGAGGCTGAGC 63088  
QY 721 AGGAGGAGCTGCTGCGCGCTGCTGAGGAGGCTGCTGAGGAGGCTGCTGAGGAGGCTGAGC 780  
DB 63089 AGGAGGAGCTGCTGCGCGCTGCTGAGGAGGCTGCTGAGGAGGCTGCTGAGGAGGCTGAGC 63148  
QY 781 TACCTTATGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
DB 63149 TACCTTATGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63208  
QY 841 AAGGCGGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
DB 63209 AAGGCGGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63268  
QY 901 CTCAGAAATGAGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
DB 63269 CTCAGAAATGAGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63328  
QY 961 AAGAAAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
DB 63329 AAGAAAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63388  
QY 1021 TCCAGGCGCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080  
DB 63389 TCCAGGCGCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63448  
QY 1081 GCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
DB 63449 GCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63508  
QY 1141 ACCATCTTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
DB 63509 ACCATCTTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63568  
QY 1201 CCGCGCTTCCGAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1260  
DB 63569 CCGCGCTTCCGAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 63628

RESULT 2  
US-09-568-102-1

; Sequence 1, Application US/09568102  
; Patent No. 6346404

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Rose

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

FILE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,102  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patemcin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-568-102-1  
Query Match 100.0%; Score 1260; DB 4; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 9,96-250;  
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACACAGGAGCAAGCGAATCAGATGAGACGAAAGCGCTTTCGACTTCAAGCGCTTC 60  
DB 62369 ATGACACAGGAGCAAGCGAATCAGATGAGAGGAGGCTTTCGACTTCAAGCGCTTC 62428  
QY 61 GCGCTGAGGTAAGCGGAGGAGCCGTTCCCGGATGAGCGGCTGAGAGGCAACCCG 120  
DB 62429 GCGCTGAGGTAAGCGGAGGAGCCGTTCCCGGATGAGCGGCTGAGAGGCAACCCG 62488  
QY 121 ATCTTCTAGGAGTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
DB 62489 ATCTTCTAGGAGTGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62548  
QY 181 GGTTCGCGGAGAAAGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
DB 62549 GGTTCGCGGAGAAAGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62608  
QY 241 TCGTGGGCAATCCGAGCTCAGCGATATGAAGATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
DB 62609 TCGTGGGCAATCCGAGCTCAGCGATATGAAGATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62668  
QY 301 GAGGATCAGGCTGCGGCTGCGCAAGCTGCTCAACCGCTGCTTACGCTCAGCGGCTTCCG 360  
DB 62669 GAGGATCAGGCTGCGGCTGCGCAAGCTGCTCAACCGCTGCTTACGCTCAGCGGCTTCCG 62728  
QY 361 CTGCTGCGCGCGGAAATACAGCGCACCGTGCAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
DB 62729 CTGCTGCGCGCGGAAATACAGCGCACCGTGCAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62788  
QY 421 GAGGAGTTGAGCGTTGTGCGGAGTATACGCGAGGGAATCCGATGCGCGGATCAGCGCT 480  
DB 62789 GAGGAGTTGAGCGTTGTGCGGAGTATACGCGAGGGAATCCGATGCGCGGATCAGCGCT 62848  
QY 481 CTGTTGAAGTTCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGGCACTGCG 540  
DB 62849 CTGTTGAAGTTCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGGCACTGCG 62908  
QY 541 CCGCGCGCTGCGCGGTTTGTGCTGCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGG 600  
DB 62909 CCGCGCGCTGCGCGGTTTGTGCTGCGCGCGAGTGTGACGAGAAAGTTCCGTCGCTTCCGCTCGG 62968  
QY 601 TCCGTCACGAGGAGGCTGCGCGGCTGCTCAATGAGTGTGACGAGGAGGCGGAGAAACCGG 660  
DB 62969 TCCGTCACGAGGAGGCTGCGCGGCTGCTCAATGAGTGTGACGAGGAGGCGGAGAAACCGG 63028  
QY 661 CTGGAAGATGACGCTTCTTGAAGATGCTGCTGAGGCGGAGGCGGAGGCGGAGGCTGAGC 720  
DB 63029 CTGGAAGATGACGCTTCTTGAAGATGCTGCTGAGGCGGAGGCGGAGGCGGAGGCTGAGC 63088  
QY 721 AGGAGGAGCTGCTGCGCGCTGCTGAGGAGGCTGCTGAGGAGGCTGCTGAGGAGGCTGAGC 780  
DB 63089 AGGAGGAGCTGCTGCGCGCTGCTGAGGAGGCTGCTGAGGAGGCTGCTGAGGAGGCTGAGC 63148  
QY 781 TACCTTATGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
DB 63149 TACCTTATGCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63208

QY 841 AAGCCGAGCCCGGCTCATGAGAAACGGCTCATGAGTGTCTCCGTTGCACATATC 900  
DB 63209 AAGCCGAGCCCGGCTCATGAGAAACGGCTCATGAGTGTCTCCGTTGCACATATC 63268  
QY 901 CTCGAATAGGAACCTGTGCTTCCGCGAGGACCTGTGAGTACTGCGGGGCATTCATC 960  
DB 63269 CTCGAATAGGAACCTGTGCTTCCGCGAGGACCTGTGAGTACTGCGGGGCATTCATC 63328  
QY 961 AAGAAAGGGAGATGCTTCTCTCTGATCCCGAGCCCTGTAGAGATGGAGCTGATTC 1020  
DB 63329 AAGAAAGGGAGATGCTTCTCTCTGATCCCGAGCCCTGTAGAGATGGAGCTGATTC 63388  
QY 1021 TCCAGGCCAGACGTGTGATGTGACGGGACACGGGCGCGAGCTCGCTGAGTATG 1080  
DB 63389 TCCAGGCCAGACGTGTGATGTGACGGGACACGGGCGCGAGCTCGCTGAGTATG 63448  
QY 1081 GGGCCCATGTCTGCCCCGGGGGTGTCTGTGCTGTGAGGCGGAGATGCGCTGGGC 1140  
DB 63449 GGGCCCATGTCTGCCCCGGGGGTGTCTGTGCTGTGAGGCGGAGATGCGCTGGGC 63508  
QY 1141 ACCATCTTCGCTAGGTTCCCGAGATGAGCTGAAGAACTCCGCTGTGGATATACAC 1200  
DB 63509 ACCATCTTCGCTAGGTTCCCGAGATGAGCTGAAGAACTCCGCTGTGGATATACAC 63568  
QY 1201 CCCGCGTTCGGAACATGCAATCACTCAACGTCACTTTGAAGCCCTCCAAAGCTGATAG 1260  
DB 63569 CCCGCGTTCGGAACATGCAATCACTCAACGTCACTTTGAAGCCCTCCAAAGCTGATAG 63628

RESULT 3  
US-09-567-969-1  
; Sequence 1, Application US/09567969  
; Patent No. 6355457

GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zierle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/567,969  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 68750  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-09-567-969-1

Query Match 100.0%; Score 1260; DB 4; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 9,9e-290;  
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACACGAGCAACGAATCAGTAGAGAGCAAGCTGCTTCAAGCCGTTG 60  
DB 62369 ATGACACGAGCAACGAATCAGTAGAGAGCAAGCTGCTTCAAGCCGTTG 62428  
QY 61 GCGCGTGGGTACGCGGAGAGACCCGTTCCCGGATCGAGCGCTGAGAGAGCAACCCG 120  
DB 62429 GCGCGTGGGTACGCGGAGAGACCCGTTCCCGGATCGAGCGCTGAGAGAGCAACCCG 62488  
QY 121 ATCTTCTACTGAGATGAGCGCTCTGTGCTCTCAACCCGATACAGAGCTGTGCGG 180  
DB 62489 ATCTTCTACTGAGATGAGCGCGCTCTGTGCTCTCAACCCGATACAGAGCTGTGCGG 62548  
QY 181 GTGTTCCGCGAAGCGCTTGGCGGTCACTGAGAAAGTGGGAATCAAGCGCGAGTAC 240

DB 62549 GTGTTCCGCGAAGCGCTTGGCGGTCACTGAGAAAGTGGGAATCAAGCGCGAGTAC 62608  
QY 241 TGTGCGCATTTCCCGAGCTCAGGATATGAAGATGCGATTGTTGCGGCTCGCGG 300  
DB 62609 TGTGCGCATTTCCCGAGCTCAGGATATGAAGATGCGATTGTTGCGGCTCGCGG 62668  
QY 301 GAGATCAACGCTCGAGTCCGAGCTGTCAACCCGTCGTTACGTCACGCGCATTCAC 360  
DB 62669 GAGATCAACGCTCGAGTCCGAGCTGTCAACCCGTCGTTACGTCACGCGCATTCAC 62728  
QY 361 CTGCTGCGCGGAAATACAGCGGACCGGTGACCAAGCTGCTGATGCTGCTCCGGAACA 420  
DB 62729 CTGCTGCGCGGAAATACAGCGGACCGGTGACCAAGCTGCTGATGCTGCTCCGGAACA 62788  
QY 421 GAGAGTTGCACTGTGTCGGGATTTACGCGGAGGAAATCCCGATGCGCGCATAGCGCT 480  
DB 62789 GAGAGTTGCACTGTGTCGGGATTTACGCGGAGGAAATCCCGATGCGCGCATAGCGCT 62848  
QY 481 CTGTTGAAGGTTCCGCGCGAGTGTACGAGAAAGTTCCGTCGCTTCGCTCGGCACTGCG 540  
DB 62849 CTGTTGAAGGTTCCGCGCGAGTGTACGAGAAAGTTCCGTCGCTTCGCTCGGCACTGCG 62908  
QY 541 CGCGCGCTCGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600  
DB 62909 CGCGCGCTCGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 62968  
QY 601 TCCGTCACGAGGAGCTCGCGCTGCTCATACGCTCTCATACGCTCTCATACGCTCTCATAC 660  
DB 62969 TCCGTCACGAGGAGCTCGCGCTGCTCATACGCTCTCATACGCTCTCATACGCTCTCATAC 63028  
QY 661 CTCGAAATGACGCTTGTGACGATGCTGTTGAGGCGGAGCGGACGAGGAGCTGAGC 720  
DB 63029 CTCGAAATGACGCTTGTGACGATGCTGTTGAGGCGGAGCGGACGAGGAGCTGAGC 63088  
QY 721 ACGAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
DB 63089 ACGAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63148  
QY 781 TACCTTATCGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
DB 63149 TACCTTATCGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63208  
QY 841 AAGCCGAGCCCGGCTCATGAGAAACGGCTCATGAGTGTCTCCGTTGCACATATC 900  
DB 63209 AAGCCGAGCCCGGCTCATGAGAAACGGCTCATGAGTGTCTCCGTTGCACATATC 63268  
QY 901 CTCGAATAGGAACCTGTGCTTCCGCGAGGACCTGTGAGTACTGCGGGGCATTCATC 960  
DB 63269 CTCGAATAGGAACCTGTGCTTCCGCGAGGACCTGTGAGTACTGCGGGGCATTCATC 63328  
QY 961 AAGAAAGGGAGATGCTTCTCTCTGATCCCGAGCCCTGTAGAGATGGAGCTGATTC 1020  
DB 63329 AAGAAAGGGAGATGCTTCTCTCTGATCCCGAGCCCTGTAGAGATGGAGCTGATTC 63388  
QY 1021 TCCAGGCCAGACGTGTGATGTGACGGGACACGGGCGCGAGCTCGCTGAGTATG 1080  
DB 63389 TCCAGGCCAGACGTGTGATGTGACGGGACACGGGCGCGAGCTCGCTGAGTATG 63448  
QY 1081 GGGCCCATGTCTGCCCCGGGGGTGTCTGTGCTGTGAGGCGGAGATGCGCTGGGC 1140  
DB 63449 GGGCCCATGTCTGCCCCGGGGGTGTCTGTGCTGTGAGGCGGAGATGCGCTGGGC 63508  
QY 1141 ACCATCTTCGCTAGGTTCCCGAGATGAGCTGAAGAACTCCGCTGTGGATATACAC 1200  
DB 63509 ACCATCTTCGCTAGGTTCCCGAGATGAGCTGAAGAACTCCGCTGTGGATATACAC 63568  
QY 1201 CCCGCGTTCGGAACATGCAATCACTCAACGTCACTTTGAAGCCCTCCAAAGCTGATAG 1260  
DB 63569 CCCGCGTTCGGAACATGCAATCACTCAACGTCACTTTGAAGCCCTCCAAAGCTGATAG 63628

RESULT 4  
US-09-568-480-1



Db 62429 GGGCTGGGTACGCGAGAGACCCGTTCCCGGATCGAGCGCTGAGAGCAACCCCC 62488  
Qy 121 ATCTTCTAGGAGATGAAAGCGCTCTGGGTCTCAACCCGATACAGAGCGTGCAGG 180  
Db 62489 ATCTTCTAGGAGATGAAAGCGCTCTGGGTCTCAACCCGATACAGAGCGTGCAGG 62548  
Qy 181 GTGTTCCGCGACGAACGCTTCGCGGTCACTCGAGAAAGTGGAAATCGAGCGGAGTAC 240  
Db 62549 GTGTTCCGCGACGAACGCTTCGCGGTCACTCGAGAAAGTGGAAATCGAGCGGAGTAC 62608  
Qy 241 TCGTGGCGCATTTCCGAGCTCAAGCATATGAAAGTACGATTTGTTGGGCTGCCGCG 300  
Db 62609 TCGTGGCGCATTTCCGAGCTCAAGCATATGAAAGTACGATTTGTTGGGCTGCCGCG 62668  
Qy 301 GAGGATCAAGCTCGGGTTCGCGAAGCTCGTCAACCCGTTACGTCACGCGCATCGAC 360  
Db 62669 GAGGATCAAGCTCGGGTTCGCGAAGCTCGTCAACCCGTTACGTCACGCGCATCGAC 62728  
Qy 361 CTGCTCGCGCGCGAAATACAGCGCACCGTCAACAGCTGCTCGATGCTCGCGACAA 420  
Db 62729 CTGCTCGCGCGCGAAATACAGCGCACCGTCAACAGCTGCTCGATGCTCGCGACAA 62788  
Qy 421 GAGGATTCGAGCTTGTGCGGATTAACGCGAGGGAATCCGATGCGCGCATCGCGCT 480  
Db 62789 GAGGATTCGAGCTTGTGCGGATTAACGCGAGGGAATCCGATGCGCGCATCGCGCT 62848  
Qy 481 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAATCCGCTGCTGGCGCTCGCGACTGC 540  
Db 62849 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAATCCGCTGCTGGCGCTCGCGACTGC 62908  
Qy 541 CGCGCGCTCGCGGTGGTGGTCCCGACAGTCAATGAGAGAACCAAGCCCTGGTCCG 600  
Db 62909 CGCGCGCTCGCGGTGGTGGTCCCGACAGTCAATGAGAGAACCAAGCCCTGGTCCG 62968  
Qy 601 TCCGTCACGAGAGGCGTCCGCTGCTCAATGAGTCTTGATGAGCGCGCAGAACCCG 660  
Db 62969 TCCGTCACGAGAGGCGTCCGCTGCTCAATGAGTCTTGATGAGCGCGCAGAACCCG 63028  
Qy 661 CTGGAATGAGCTGCTGAGTCAAGTCTGCTCAAGCGCGAGCGCGAGCGCGCTGAGC 720  
Db 63029 CTGGAATGAGCTGCTGAGTCAAGTCTGCTCAAGCGCGAGCGCGAGCGCGCTGAGC 63088  
Qy 721 ACGAAGAGCTGCTCGCGCTCGTGGGTGATTAATGCTGCGACCGATACCAAGTAC 780  
Db 63089 ACGAAGAGCTGCTCGCGCTCGTGGGTGATTAATGCTGCGACCGATACCAAGTAC 63148  
Qy 781 TACCTTATCGCTGCTGCTCAACCTGCTGCGGTGCGCGAGCGCTCGAGCTGCTG 840  
Db 63149 TACCTTATCGCTGCTGCTCAACCTGCTGCGGTGCGCGAGCGCTCGAGCTGCTG 63208  
Qy 841 AAGGCGAGCGCGCGCTCAATGAGAAACCGGCTCGATGAGTCTCGCTTCACAATATC 900  
Db 63209 AAGGCGAGCGCGCGCTCAATGAGAAACCGGCTCGATGAGTCTCGCTTCACAATATC 63268  
Qy 901 CTGCAATAGAACTGTGCGTTCGCGACAGGACGAACTGAGTACTGCGGGCATCGATC 960  
Db 63269 CTGCAATAGAACTGTGCGTTCGCGACAGGACGAACTGAGTACTGCGGGCATCGATC 63328  
Qy 961 AAGAAAGGAGATGCTTCTTCTCTGATCCGAGCGCTCGAGAGATGGAGCTGTATTC 1020  
Db 63329 AAGAAAGGAGATGCTTCTTCTCTGATCCGAGCGCTCGAGAGATGGAGCTGTATTC 63388  
Qy 1021 TCCAGGCGAGAGTGTGATGATGCGACGCGGACAGCGGCGCGACGCTCGCGATGATA 1080  
Db 63389 TCCAGGCGAGAGTGTGATGATGCGACGCGGACAGCGGCGCGACGCTCGCGATGATA 63448  
Qy 1081 GCGCGCGCATGTGCGCGCGGAGTGTCCCTGCTCGCTCGAGCGGAGATCCCGCTGGC 1140  
Db 63449 GCGCGCGCATGTGCGCGCGGAGTGTCCCTGCTCGCTCGAGCGGAGATCCCGCTGGC 63508  
Qy 1141 ACCATCTTCGTAAGTTCCTCGAGATGAAAGTAAAGAACTCCGCTGTTGGATACAC 1200

Db 63509 ACCATCTTCGTAAGTTCCTCGAGATGAAAGTAAAGAACTCCGCTGTTGGATACAC 63568  
Qy 1201 CCGCGCTCCGGAACATGATCACTCAACGCTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260  
Db 63569 CCGCGCTCCGGAACATGATCACTCAACGCTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628  
  
RESULT 6  
US-09-568-472-1  
; Sequence 1, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-10582A  
; CURRENT APPLICATION NUMBER: US/09/568,472  
; PRIOR FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-568-472-1  
  
Query Match 100.0%; Score 1260; DB 4; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 9.9e-280;  
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 ATGACACAGAGCAACGAAATCAGATGAGACGAAGCTGCTTGAAGCTTCAAGCCGTT 60  
Db 62369 ATGACACAGAGCAACGAAATCAGATGAGACGAAGCTGCTTGAAGCTTCAAGCCGTT 62428  
Qy 61 GCGCTGGGTACCGGAGAGACCCGTTCCCGGATGAGCGCTGAGAGAGCAACCC 120  
Db 62429 GCGCTGGGTACCGGAGAGACCCGTTCCCGGATGAGCGCTGAGAGAGCAACCC 62488  
Qy 121 ATCTTCTAGGAGATGAAAGCGCTCTGGGTCTCAACCCGATACAGAGCGTGCAGG 180  
Db 62489 ATCTTCTAGGAGATGAAAGCGCTCTGGGTCTCAACCCGATACAGAGCGTGCAGG 62548  
Qy 181 GTGTTCCGCGACGAACGCTTCGCGGTCACTCGAGAAAGTGGAAATCGAGCGGAGTAC 240  
Db 62549 GTGTTCCGCGACGAACGCTTCGCGGTCACTCGAGAAAGTGGAAATCGAGCGGAGTAC 62608  
Qy 241 TCGTGGCGCATTTCCGAGCTCAAGCATATGAAAGTACGATTTGTTGGGCTGCCGCG 300  
Db 62609 TCGTGGCGCATTTCCGAGCTCAAGCATATGAAAGTACGATTTGTTGGGCTGCCGCG 62668  
Qy 301 GAGGATCAAGCTCGGGTTCGCGAAGCTCGTCAACCCGTTACGTCACGCGCATCGAC 360  
Db 62669 GAGGATCAAGCTCGGGTTCGCGAAGCTCGTCAACCCGTTACGTCACGCGCATCGAC 62728  
Qy 361 CTGCTCGCGCGCGAAATACAGCGCACCGTCAACAGCTGCTCGATGCTCGCGACAA 420  
Db 62729 CTGCTCGCGCGCGAAATACAGCGCACCGTCAACAGCTGCTCGATGCTCGCGACAA 62788  
Qy 421 GAGGATTCGAGCTTGTGCGGATTAACGCGAGGGAATCCGATGCGCGCATCGCGCT 480  
Db 62789 GAGGATTCGAGCTTGTGCGGATTAACGCGAGGGAATCCGATGCGCGCATCGCGCT 62848  
Qy 481 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAATCCGCTGCTGGCGCTCGCGACTGC 540  
Db 62849 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAATCCGCTGCTGGCGCTCGCGACTGC 62908  
Qy 541 CGCGCGCTCGCGGTGGTGGTCCCGACAGTCAATGAGAGAACCAAGCCCTGGTCCG 600

```

Db      62909  CGGCGCTCGCGTGGTGTGGCCCGCAGGTGATGAGAGCAAGACCTGTGGCG 62968
Qy      601     TCCGTCACCGAGGGGCTCCGCGTCTCATGACGTCTCGATGAGCGGGCAGAAACCCG 660
Db      62969  TCCGTCACCGAGGGGCTCCGCGTCTCATGACGTCTCGATGAGCGGGCAGAAACCCG 63028
Qy      661     CTCGAAATGACGTCTTGAACGATGCTGCTTCAGGCGGAGCGGACAGCGCTGAGC 720
Db      63029  CTCGAAATGACGTCTTGAACGATGCTGCTTCAGGCGGAGCGGACAGCGCTGAGC 63088
Qy      721     ACGAAGAGCTGCTGCGCTCGGAGGAGGATTAATGCTGCTGAGCAACGATACAGATC 780
Db      63089  ACGAAGAGCTGCTGCGCTCGGAGGAGGATTAATGCTGCTGAGCAACGATACAGATC 63148
Qy      781     TACCTTATGCGCTTCCGCTGCTGCTCAACCTGCTGCGGCTCGGAGCGCTGAGCTGCTG 840
Db      63149  TACCTTATGCGCTTCCGCTGCTGCTCAACCTGCTGCGGCTCGGAGCGCTGAGCTGCTG 63208
Qy      841     AAGGCGGAGCGCGGCTCATGAGAAACGGCTGATGAGGTGCTCCGCTTCGACATATTC 900
Db      63209  AAGGCGGAGCGCGGCTCATGAGAAACGGCTGATGAGGTGCTCCGCTTCGACATATTC 63268
Qy      901     CTCAGAAATGAGAACTGTGGCTTTCGCGAGGAGGAGCTGAGATCTGCGGGGATGATTC 960
Db      63269  CTCAGAAATGAGAACTGTGGCTTTCGCGAGGAGGAGCTGAGATCTGCGGGGATGATTC 63328
Qy      961     AAGAAAGGGAGATGTCTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 1020
Db      63329  AAGAAAGGGAGATGTCTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 63388
Qy      1021  TCCAGGCCAGACGTGTTTGTATGTGCGACGGGACACGGGCGGAGCTCCGCTACGCTATG 1080
Db      63389  TCCAGGCCAGACGTGTTTGTATGTGCGACGGGACACGGGCGGAGCTCCGCTACGCTATG 63448
Qy      1081  GGGCCCCATGTCTGCCCCGGGGGTGCTCCCTTGTGCTGCTGAGGCGGAGATTCGCGCGG 1140
Db      63449  GGGCCCCATGTCTGCCCCGGGGGTGCTCCCTTGTGCTGCTGAGGCGGAGATTCGCGCGG 63508
Qy      1141  ACCATCTTCGTAGTGTCCCGAGATGAAAGCTGAAAGAACTCCGCTGTTTGGATACAC 1200
Db      63509  ACCATCTTCGTAGTGTCCCGAGATGAAAGCTGAAAGAACTCCGCTGTTTGGATACAC 63568
Qy      1201  CCGCGCTTCGGAACATCGAATACCTGATCTTGAAGCCCTTCGAAAGCTGATG 1260
Db      63569  CCGCGCTTCGGAACATCGAATACCTGATCTTGAAGCCCTTCGAAAGCTGATG 63628

RESULT 7
US-09-567-899-1
/ Sequence 1, Application US/09567899
/ Patent No. 6383787
/ GENERAL INFORMATION:
/ APPLICANT: Schupp, Thomas
/ APPLICANT: Ligon, James
/ APPLICANT: Molnar, Istvan
/ APPLICANT: Zirkle, Ross
/ APPLICANT: Cyr, Devon
/ APPLICANT: Goetlach, Joern
/ TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
/ FILE REFERENCE: 4-30582A
/ CURRENT APPLICATION NUMBER: US/09/567, 899
/ PRIOR FILING DATE: 2000-05-10
/ PRIOR FILING DATE: 1999-06-17
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 68750
/ TYPE: DNA
/ ORGANISM: Sorangium cellulosum
US-09-567-899-1

```

```

Query Match      100.0%; Score 1260; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9.9e-250;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1     ATGACACAGAGCAACGCAATCAGATGAGAGCAAGCTGTCTTTCGACTTCAAGCGCTTC 60
Db      62369  ATGACACAGAGCAACGCAATCAGATGAGAGCAAGCTGTCTTTCGACTTCAAGCGCTTC 62428
Qy      61     GCGCTGGGATACCGGAGAGACCGTTCCCGGATGAGCGCTCTGAGAGAGCAACCCG 120
Db      62429  GCGCTGGGATACCGGAGAGACCGTTCCCGGATGAGCGCTCTGAGAGAGCAACCCG 62488
Qy      121  ATCTTCTACTGGGATGAGAGCGGCTCTGGGCTCACCGGATACACAGAGCTGTGCGG 180
Db      62489  ATCTTCTACTGGGATGAGAGCGGCTCTGGGCTCTCACCGGATACACAGAGCTGTGCGG 62548
Qy      181  GTGTTCCGCGACGAAACGCTTCGCGTCACTGAGAGAGTGGAAATCGAGCGGAGTAC 240
Db      62549  GTGTTCCGCGACGAAACGCTTCGCGTCACTGAGAGAGTGGAAATCGAGCGGAGTAC 62608
Qy      241  TCGTGGCCATTCGCGAGCTCAGCGATGAAAGATACGATTGTTGGGCTGCGCGG 300
Db      62609  TCGTGGCCATTCGCGAGCTCAGCGATGAAAGATACGATTGTTGGGCTGCGCGG 62668
Qy      301  GAGGATCAACGCTCGGCTCGGAGCTCGCAAGCTCGTCAACCGGTGTTAGCTACGCGCATTCGAC 360
Db      62669  GAGGATCAACGCTCGGCTCGGAGCTCGCAAGCTCGTCAACCGGTGTTAGCTACGCGCATTCGAC 62728
Qy      361  CTGCTCGCGCGCAAAATCAGCGCACCGTTCAGACAGCTGCTGATGCTCGCTCCGAGCA 420
Db      62729  CTGCTCGCGCGCAAAATCAGCGCACCGTTCAGACAGCTGCTGATGCTCGCTCCGAGCA 62788
Qy      421  GAGGATTCGACGTTGTGCGGATTAACGGAGGGAATCCGATCCGCGCATTCAGCGCT 480
Db      62789  GAGGATTCGACGTTGTGCGGATTAACGGAGGGAATCCGATCCGCGCATTCAGCGCT 62848
Qy      481  CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAACTCCGCTCGTTCGCGCTCGGAGCTGCG 540
Db      62849  CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAACTCCGCTCGTTCGCGCTCGGAGCTGCG 62908
Qy      541  CCGCGCTCGCGGCTGTTGTGCTCCGAGGTGATGAGAGACCAAGACCTGTGTGCG 600
Db      62909  CCGCGCTCGCGGCTGTTGTGCTCCGAGGTGATGAGAGACCAAGACCTGTGTGCG 62968
Qy      601  TCCGTCACCGAGGCGCTGCGCTGCTCAATGACGTCTGATGAGCGGCGGAGAAACCG 660
Db      62969  TCCGTCACCGAGGCGCTGCGCTGCTCAATGACGTCTGATGAGCGGCGGAGAAACCG 63028
Qy      661  CTCGAAATGACGTCTTGAACGATGCTGCTTCAGGCGGAGCGGACAGCGCTGAGC 720
Db      63029  CTCGAAATGACGTCTTGAACGATGCTGCTTCAGGCGGAGCGGACAGCGCTGAGC 63088
Qy      721  ACGAAGAGCTGCTGCGCTCGGAGGAGGATTAATGCTGCTGAGCAACGATACAGATC 780
Db      63089  ACGAAGAGCTGCTGCGCTCGGAGGAGGATTAATGCTGCTGAGCAACGATACAGATC 63148
Qy      781  TACCTTATGCGCTTCCGCTGCTGCTCAACCTGCTGCGGCTCGGAGCGCTCAGGCTGAG 840
Db      63149  TACCTTATGCGCTTCCGCTGCTGCTCAACCTGCTGCGGCTCGGAGCGCTCAGGCTGAG 63208
Qy      841  AAGGCGGAGCGCGGCTCATGAGAAACGGCTGATGAGGTGCTCGGCTTCGACATATTC 900
Db      63209  AAGGCGGAGCGCGGCTCATGAGAAACGGCTGATGAGGTGCTCGGCTTCGACATATTC 63268
Qy      901  CTCAGAAATGAGAACTGTGGCTTTCGCGAGGAGGAGCTGAGTACTGCGGGCATGATC 960
Db      63269  CTCAGAAATGAGAACTGTGGCTTTCGCGAGGAGGAGCTGAGTACTGCGGGCATGATC 63328
Qy      961  AAGAAAGGGAGATGTCTTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 1020
Db      63329  AAGAAAGGGAGATGTCTTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 63388
Qy      1021  TCCAGGCCAGACGTGTTTGTATGTGCGACGGGACACGGGCGGAGCTCGCTGACGCTATG 1080

```

Db 63389 TCCAGGCCAGACGTTTGTGAGACGAGCAGGCGCAGCTGCGTACCGTGA 63448  
Qy 1081 GGGCCCCCATGCTGTCGCCCGGGGGTCCCTGTGCTGCGCTCGAGGGGAGATGCGGTGAGC 1140  
Db 63449 GGGCCCCCATGCTGTCGCCCGGGGGTCCCTGTGCTGCGCTCGAGGGGAGATGCGGTGAGC 63508  
Qy 1141 ACCATCTTCGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATACAC 1200  
Db 63509 ACCATCTTCGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATACAC 63568  
Qy 1201 CCGCGCTTCGGAACATGGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260  
Db 63569 CCGCGCTTCGGAACATGGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 8  
US-09-443-501A-2  
; Sequence 2, Application US/09443501A  
; Patent No. 6303342  
; GENERAL INFORMATION:  
; APPLICANT: Kosan Biosciences, Inc.  
; APPLICANT: Julien, Bryan  
; APPLICANT: Katz, Leonard  
; APPLICANT: Khosla, Chaitan  
; APPLICANT: Tang, Li  
; APPLICANT: Ziermann, Rainer  
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing  
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives  
; FILE REFERENCE: 30062-20031.00  
; CURRENT FILING DATE: US/09/443, 501A  
; PRIOR FILING DATE: 1999-11-19  
; PRIOR APPLICATION NUMBER: US 60/130,560  
; PRIOR FILING DATE: 1999-04-22  
; PRIOR APPLICATION NUMBER: US 60/122,620  
; PRIOR FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: US 60/119,386  
; PRIOR FILING DATE: 1999-02-10  
; PRIOR APPLICATION NUMBER: US 60/109,401  
; PRIOR FILING DATE: 1998-11-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 71989  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic construct  
; US-09-443-501A-2

Query Match 98.9% Score 1246.2; DB 4; Length 71989;  
Best Local Similarity 99.4%; Pred. No. 1.9e-286;  
Matches 1251; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGACACAGAGCAAGCAATCAGAGTGAAGCAAGCTGCTTTCGACTTCAAGCCGTTTC 60  
Db 56757 ATGACACAGAGCAAGCAATCAGAGTGAAGCAAGCTGCTTTCGACTTCAAGCCGTTTC 56816  
Qy 61 GCGCTGGGTACGCGGAGAGCCGTTCCCGCGATCGAGCGCTGAGAGGCAACCCCC 120  
Db 56817 GCGCTGGGTACGCGGAGAGCCGTTCCCGCGATCGAGCGCTGAGAGGCAACCCCC 56876  
Qy 121 ATCTTCTACGAGATGAAGCGCGCTCGGGTCTCAACCCGATACAGAGTGTGCGGCG 180  
Db 56877 ATCTTCTACGAGATGAAGCGCGCTCGGGTCTCAACCCGATACAGAGTGTGCGGCG 56936  
Qy 181 GTGTCCCGCAGCAAGCTTTCGCGCTCAGTCGAGAGAGTGGGATTCGAGCGCGAGTAC 240  
Db 56937 GTGTCCCGCAGCAAGCTTTCGCGCTCAGTCGAGAGAGTGGGATTCGAGCGCGAGTAC 56996  
Qy 241 TCGTCCGCAATTCGCGAGCTCAGCATATGAAGAAATGCGATTTGTTGGCGTCCGCGG 300  
Db 56997 TCGTCCGCAATTCGCGAGCTCAGCATATGAAGAAATGCGATTTGTTGGCGTCCGCGG 57056

Qy 301 GAGATCAAGCTCGGATCCGCAAGCTGTCACACCCGTCGTTTACATCAGCGCCATCGAC 360  
Db 57057 GAGATCAAGCTCGGATCCGCAAGCTGTCACACCCGTCGTTTACATCAGCGCCATCGAC 57116  
Qy 361 CTGCTGGCGCGCAATACAGCCGACCGTCAACAGCTGCTGATGCTGCTCCGAGCA 420  
Db 57117 CTGCTGGCGCGCAATACAGCCGACCGTCAACAGCTGCTGATGCTGCTCCGAGCA 57176  
Qy 421 GAGAGTTGACGTTGTGCGGAGTTACCGGAGGAAATCCGATGCGCGCATCGAGCT 480  
Db 57177 GAGAGTTGACGTTGTGCGGAGTTACCGGAGGAAATCCGATGCGCGCATCGAGCT 57236  
Qy 481 CTGTTGAAGTTCCGCGCGAGTGTGACGAGAAAGTTCCGCTTCGCTTCGCGCATGCG 540  
Db 57237 CTGTTGAAGTTCCGCGCGAGTGTGACGAGAAAGTTCCGCTTCGCTTCGCGCATGCG 57296  
Qy 541 CCGCGCTCGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600  
Db 57297 CCGCGCTCGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660  
Qy 601 TCCGTCACCGAGGGGCTCGCGCTGCTCATGACGTCCTCGATGAGCGCGCGAGAACCG 660  
Db 57357 TCCGTCACCGAGGGGCTCGCGCTGCTCATGACGTCCTCGATGAGCGCGCGAGAACCG 57416  
Qy 661 CTCGAAATGACGCTTTCGACGATGCTGCTTCAGGCGCGAGCGCGAGCGAGCTGAGC 720  
Db 57417 CTCGAAATGACGCTTTCGACGATGCTGCTTCAGGCGCGAGCGCGAGCGAGCTGAGC 57476  
Qy 721 ACGAAGAGCTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 780  
Db 57477 ACGAAGAGCTGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 57536  
Qy 781 TACCTTATCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
Db 57537 TACCTTATCGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57596  
Qy 841 AAGCCGAGCCCGGCTCATGAGAAAGCGCTGATGAGAGTGTGCTGCTGCTGCTGCTGCT 900  
Db 57597 AAGCCGAGCCCGGCTCATGAGAAAGCGCTGATGAGAGTGTGCTGCTGCTGCTGCTGCT 57656  
Qy 901 CTCGAATAGGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 57657 CTCGAATAGGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57716  
Qy 961 AAGAAAGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
Db 57717 AAGAAAGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57776  
Qy 1021 TCCAGGCCAGAGCTGTTGATGTCGAGCGGAGCAAGCGGCGGAGCTGCTGCTGCTGCTGCT 1080  
Db 57777 TCCAGGCCAGAGCTGTTGATGTCGAGCGGAGCAAGCGGCGGAGCTGCTGCTGCTGCTGCT 57836  
Qy 1081 GGGCCCCCATGCTGTCGCCCGGGGGTCCCTGTGCTGCGCTCGAGGGGAGATGCGGTGAGC 1140  
Db 57837 GGGCCCCCATGCTGTCGCCCGGGGGTCCCTGTGCTGCGCTCGAGGGGAGATGCGGTGAGC 57896  
Qy 1141 ACCATCTTCGTAGGTTCCCGAGATGAAGCTGAAGAAATCCTCCGTTGTTGATACAC 1200  
Db 57897 ACCATCTTCGTAGGTTCCCGAGATGAAGCTGAAGAAATCCTCCGTTGTTGATACAC 57956  
Qy 1201 CCGCGCTTCGGAACATGGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1259  
Db 57957 CCGCGCTTCGGAACATGGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 58015

RESULT 9  
US-09-252-991A-1265  
; Sequence 1265, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 1265  
LENGTH: 1314  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1265

Query Match 11.5%; Score 145.4; DB 4; Length 1314;  
Best Local Similarity 48.8%; Pred. No. 1.1e-25;  
Matches 456; Conservative 0; Mismatches 446; Indels 33; Gaps 1;

254 CCGAGCTCAGCATATGAAAGATGAGATTGTCGGGCTGCCCGCGGAGATCAGCTC 313  
DB CCGGATCGCGGAGGAGGAGTCAACATGCTCACTCGACCGCGGACATACCC 349  
314 GGGTCGCAAGCTCGTCAACCGCGTTCAGTCAAGCCGATCGACTGCTGCGCGCG 373  
DB GCGTGGCTCGCTGCTGCTGCGCGCTTCAACCGCGGAGTGAAGCGCTGCAACCGC 409  
350 GCGTGGCTCGCTGCTGCTGCGCGCTTCAACCGCGGAGTGAAGCGCTGCAACCGC 409  
374 AATACAGGCAACCGTCAACGAGTCTGATGCTGCTCGGACCAAGAGAGTTGAGC 433  
DB AATATGAAACGATCAACGAGATTTGCTGAGCCATGAGCCGCGGACGCGGACGAC 469  
434 TTGTGGGATATACGGGAGGAGATCCGATGCGCGGATCAGCGCTGTTGAAGTTC 493  
DB TATAGCCGATTCGATGATCCGCTGATCAATGCGGATCTTGAAGTTCGAGCATTC 529  
470 TATAGCCGATTCGATGATCCGCTGATCAATGCGGATCTTGAAGTTCGAGCATTC 529  
494 CCGCCAGTGTGACGAGAGTTCCGTCGCTCGGCTCGGACCTGCGCGCGCTCGCGC 553  
DB CCGAGCGGAGCGGAGCAACGCGCGCATCTG----- 563  
554 TGGGTTTGGTCCCGGATGATGAGAGACCAAGACCTGTCGCTGCTGATCAGGAG 613  
DB TGGGTTTGGTCCCGGATGATGAGAGACCAAGACCTGTCGCTGCTGATCAGGAG 613  
564 -----GGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 616  
614 GCGTGGCTGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 673  
DB CCGAGGTCATCACTGCGCGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 676  
617 CCGAGGTCATCACTGCGCGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 676  
674 TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 733  
DB TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 736  
677 TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 736  
734 TCGCGCTGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793  
DB TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 796  
737 TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 796  
794 TCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 853  
DB AGCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856  
797 AGCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856  
854 GCGTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 913  
DB AACTCTGCGCAACGCGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 916  
857 AACTCTGCGCAACGCGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 916  
914 CTGTGCTGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 973  
DB TGTGTCGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 976  
\*917 TGTGTCGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 976  
974 TGTGTCGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1033  
DB AACTCTGCTGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1036  
977 AACTCTGCTGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1036  
1034 TGTGTCGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1093  
DB TGTGTCGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1093

DB 1037 GCGTCAGCTACCCGCAACCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1096  
QY 1094 GCGTCAGCTACCCGCAACCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1153  
DB 1097 GCGTCAGCTACCCGCAACCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1156  
QY 1154 GCGTCAGCTACCCGCAACCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1188  
DB 1157 GCGTCAGCTACCCGCAACCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1191

## RESULT 10

US-09-252-991A-1338/c  
Sequence 1338 Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 1338  
LENGTH: 1482  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1338

Query Match 11.5%; Score 145.4; DB 4; Length 1482;  
Best Local Similarity 48.8%; Pred. No. 1.2e-25;  
Matches 456; Conservative 0; Mismatches 446; Indels 33; Gaps 1;

254 CCGAGCTCAGCATATGAAAGATGAGATTGTCGGGCTGCCCGCGGAGATCAGCTC 313  
DB CCGGATCGCGGAGGAGGAGTCAACATGCTCACTCGACCGCGGACATACCC 1158  
1217 CCGGATCGCGGAGGAGGAGTCAACATGCTCACTCGACCGCGGACATACCC 1158  
QY 314 GGGTCGCAAGCTCGTCAACCGCGTTCAGTCAAGCCGATCGACTGCTGCGCGCG 373  
DB GCGTGGCTCGCTGCTGCTGCGCGCTTCAACCGCGGAGTGAAGCGCTGCAACCGC 1098  
1157 GCGTGGCTCGCTGCTGCTGCGCGCTTCAACCGCGGAGTGAAGCGCTGCAACCGC 1098  
374 AATACAGGCAACCGTCAACGAGTCTGATGCTGCTCGGACCAAGAGAGTTGAGC 433  
DB AATATGAAACGATCAACGAGATTTGCTGAGCCATGAGCCGCGGACGCGGACGAC 1038  
1097 AATATGAAACGATCAACGAGATTTGCTGAGCCATGAGCCGCGGACGCGGACGAC 1038  
434 TTGTGGGATATACGGGAGGAGATCCGATGCGCGGATCAGCGCTGTTGAAGTTC 493  
DB TGTGTCGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 978  
1037 TGTGTCGCTTCCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 978  
494 CCGCCAGTGTGACGAGAGTTCCGTCGCTCGGCTCGGACCTGCGCGCGCTCGCGC 553  
DB CCGAGGTCATCACTGCGCGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 944  
977 CCGAGGTCATCACTGCGCGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 944  
554 TGGGTTTGGTCCCGGATGATGAGAGACCAAGACCTGTCGCTGCTGATCAGGAG 613  
DB TGGGTTTGGTCCCGGATGATGAGAGACCAAGACCTGTCGCTGCTGATCAGGAG 613  
943 -----GGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 891  
614 GCGTGGCTGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 891  
890 CCGAGGTCATCACTGCGCGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 831  
DB TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 733  
674 TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 733  
DB TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 771  
830 TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 771  
QY 734 TCGCGCTGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793  
DB TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 711  
770 TCTTTCAGATGCTGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 711



```

Oy 794 TCGCTGCTCAACCTGCTGCGGTGCGCCGAGGCGCTCGAGCTGTGAAGCCGAGCCG 853
Db 710 ACGGCTGTGTCAACCTGCTGTGTCAACCCGAGCACTGAGCTGTGCGGGGCGAGCCGG 651
Oy 854 GGCCTCATAGAGAACCGCGCTCATAGAGTGTCTCCGCTTGAACAATATCTCTAGAAATAGAA 913
Db 650 AACTCTGCGCCCAACCGCATGAGAACTGTGTCGCGCAAGACGCCGGTGTGCGCGCTCGA 591
Oy 914 CTGTGCGTTTGGCCGAGGAGCACTGAGATACCTGCGGGGCAATCATCAAGAAAGGGAGAG 973
Db 590 TGTTCGCTTCAACCTGTGAAGACGTGAATGGAACGGGGTCAACATTTCCCGCGGCAAT 531
Oy 974 TGTGTTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAATGTATTTCTCAAGGCCAGAG 1033
Db 530 ACATCTGTGTCTCAACCTGACCGCCCAACAGAGCGCGAGCGTTGACAGATCCGAGCC 471
Oy 1034 TGTTCATGTGAGAGGAGACAGCGGCGGAGCTGTGCGTACGTAGAGAGCCCGCATGTCT 1093
Db 470 GCTTCGACCTCAACCGCAACACCGATGCGCATCTCGCTACGCGCTTCGCGCTGCACTACT 411
Oy 1094 GCCCGGGGGGTGCTGCTGCTGCGCTCGAGGGGAGATGCGGTGGGCAACATTTCCGTA 1153
Db 410 GCGTGGGCGCTTCGCTGCGCGCGGCTGAGGGGCGAGATCGCAATCCAGCGCTGCTGCGCC 351
Oy 1154 GGTTCGCCGAGATGAAGCTGAAGAACTCCCGTG 1188
Db 350 GCTTCGCCGACCTCCAGTTGGCGGTGCCCAAGCG 316

```

```

RESULT 11
US-09-252-991A-1222
; Sequence 1222, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1222
; LENGTH: 1671
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1222

```

```

Query Match 11.5%; Score 145.4; DB 4; Length 1671;
Best Local Similarity 48.8%; Pred. No. 1.2e-25;
Matches 456; Conservative 0; Mismatches 446; Indels 33; Gaps 1;

```

```

Oy 254 CCGAGCTCAGCGATATGAAGAATGATGATGTTGCGGCGCGCGAGAGATCAAGCTC 313
Db 225 CCGGATGCGGAGGAGATCTACCCACCACTGCTCACTGACCCCGGAGCAATACC 284
Oy 314 GGGTCCGCAAGCTGTCAACCGGTGTTAGCTCAAGCGCATCACTGTGCGCGCG 373
Db 285 GCTGCGCTGCTGCTGCGCGCGCGGTTCACCGCGCGCAAGTGAAGCGCTGCAACGCG 344
Oy 374 AATATACGCGACCGTCAACGCTGCTGCTGCTGCTGCGAGACAAAGAGTTGACG 433
Db 345 AATATAGACGATCAACGAGAAATGCTGAGACCGCGCGCGCAACAGCGCGAC 404
Oy 434 TTGTGCGGATTTACCGGAGGAAATCCGATGCGCGCGCATCAGCGCTGTGTAAGTTG 493
Db 405 TGAATGCGGACCTTCCGATCCCGCTGACCATGCGGTATCTTCAGAGTGTGCGGATTC 464
Oy 494 CGGCGAGTGTGACGAGAAGTTCCGTGCTTCGCGTGGCGAATGCGCGCGCTGCGCG 553

```

```

Db 465 CCGAGCGCGAGCGCAACACGCGCCGCAATCTCTG----- 498
Oy 554 TGGGTTTGTGTCGCCAGGTGATGAGAGACCAAGACCTGTGCGCTGCTCAACCGAG 613
Db 499 -----GAGCGCGAGCGGAACTGCTGTGCGCGAGAGAGGCCAGGCTTGGCGCATG 551
Oy 614 GGTTCGCGCTGCTCAATGACGTCTCGATGAGAGCGCGGAGAACCCGCTGCAAAATGACG 673
Db 552 CGCAGGTGATTAATCTGCGCGGTGCTGCTGAGAGCGCAAGCGCGGAGCGACCGAGCAG 611
Oy 674 TCTTGAAGATCTGCTTTCAGGCGCGAGCGCGAGCGAGCTGAGACAGAAAGAGCTG 733
Db 612 TCTACAGCGGCGCTGTGTGAGCGCGCGCGAGCGAGCGCGCAACTGAGCGAGCGGAACTG 671
Oy 734 TCGGCTGTGAGGTGATTAATGCTGCTGCTGCGCGAGTACAGCATTAATCTTATGCGGT 793
Db 672 TCTCATAGCGCCACCTGTGTATGATGAGCGGCTTGAAGACCACTGAACATGATGCGCA 731
Oy 794 TCGCTGCTCAACCTGCTGCGGTGCGCGCGAGCGCTGAGCTGTGAGAGCGCGAGCCG 853
Db 732 ACGGCTGTGTCAACCTGTGTGTCAACCGGAGCAATGGCGTTGCTGCGGGGCGAGCGCG 791
Oy 854 GGTTCATAGAGAACGCGCTCATAGAGTGTCTCCGCTTGAACAATATCTCTAGAAATAGAA 913
Db 792 AACTCTGCGCCCAACGCGCATGAGAACTGTCTCCGCAAGAGCGCGGCTGCGCGCTGCA 851
Oy 914 CTGTGCGTTTGGCAGGAGGACCTGGAGTACTGTGCGGGGCAATCATCAAGAAAGGGAGAG 973
Db 852 TGTTCGCTTACCGTGAAGACGTGAATGGAACGGGGTCACTATTTCCCGCGGCAAT 911
Oy 974 TGTTCCTTCTGATCTGATCCGAGCGCGCTGAGAGATGAGACTGTATTTCTCAAGCGCAGAG 1033
Db 912 ACATCTGTGTCTCAACCTGACCGCAACAGCAACAGAGCGGCTTGAACATCTCCGAGCC 971
Oy 1034 TGTTCATGTGAGAGGAGACAGCGGCGGAGCTGCTGCTGATGAGAGGCGCCCATGTCT 1093
Db 972 GCTTCGACCTCAACCGCAACCGCATGCGCATCTCGCTACGCGCTTCGCGCTGCACTACT 1031
Oy 1094 GCCCGGGGGTGTCCCTTGTGCTGCGCTCGAGGGGAGATGCGGTGGGCAACATTTCCGTA 1153
Db 1032 GCGTGGGCGCTTCGCTGCGCGCGGCTGAGGGGCGAGTGGCATTCAGCGCTGCTGCGCG 1091
Oy 1154 GGTTCGCCGAGATGAAGCTGAAGAACTCCCGTG 1188
Db 1092 GCTTCGCCGACCTCCAGTTGGCGGTGCCCAAGCG 1126

```

```

RESULT 12
US-09-105-537-38
; Sequence 38, Application US/09105537A
; Patent No. 6265302
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-38

```

```

Query Match 9.8%; Score 123; DB 3; Length 1251;
Best Local Similarity 47.5%; Pred. No. 2.3e-20;
Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;
Oy 236 AGTACTGTGCGCATTCCTCCAGCTCAGCGATATGAAGAGTACGATTTGTCGGGCTGC 295

```



Qy 896 ATATCTCAGAAATAGAACTGTGCTTTCGCCAGGACGAGACTGTGAGTACTGCGGGCAT 955  
Db 2218 GCCCGGTGAATTCGCGACCTACCGCTTCCGGGTGAGGCGGTGACCTGAGCGACGG 2277  
Qy 956 CGATCAAGAAAGGAGATGCTTCTTCTGATCCCGAGCGCCTGAGATGAGGACTG 1015  
Db 2278 TCATCTCCGCGGTGACAGGCTCTGCTGCTGCTGCGAGCGCCACCGCACCCCGGAGC 2337  
Qy 1016 TATTTCCAGGCGCAGCGTGTGATGTGCGACGGGACACGGGCGCGAGCCTCGGTACG 1075  
Db 2338 GCTTCCCGAGACCGGACCGCTTTCGACATCCGCGGAGACCGCGCCGACATCTCGCTTCG 2397  
Qy 1076 GTAGAGCCCCCATCTGCTGCCCCGGGAGTGTCTTCCTGCTCGAGCGGAGATGCGCG 1135  
Db 2398 GCCACGGGATCCACTTTCGATCTGCGCGCCCTTGGCGGTTGAGGCGCGGATCGCGC 2457  
Qy 1136 TGGGACCATCTTCTCGTGTGCTTCCCGAGATGAGTGAAGAACTCCCG 1186  
Db 2458 TCCGCGCCTTCTCGAACGCTGCGGACCTGCGCTGAGACGTCTTCCCGC 2508

## RESULT 14

US-09-141-908-11  
Sequence 11, Application US/09141908  
Patent No. 6503741  
GENERAL INFORMATION:  
APPLICANT: ASHLEY, Gary  
APPLICANT: BETLACH, Mary  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a  
FILE REFERENCE: 300622002100  
CURRENT APPLICATION NUMBER: US/09/141,908  
EARLIER FILING DATE: 1998-08-28  
EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
EARLIER FILING DATE: 1998-05-06  
EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
EARLIER FILING DATE: 1997-04-30  
EARLIER APPLICATION NUMBER: PROV. 60/076,919  
EARLIER FILING DATE: 1998-03-05  
EARLIER APPLICATION NUMBER: PROV. 60/087,080  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 11  
LENGTH: 5970  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-141-908-11

Query Match 9.8%; Score 123; DB 4; Length 5970;  
Best Local Similarity 47.5%; Pred. No. 3.3e-20;  
Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;  
Qy 236 AGTACTGTGCGGCAATCCCGAGCTCAGGATATGAAGAAGTATGTTGCGGGCTGC 295  
Db 1582 ACTCAGCACTCTCCCTGACCGAGGCGCGGCGCTCAACCAACATGCTGGAGTCCG 1641  
Qy 296 CGCCGAGATCACTCTCGGCTCCGCAAGCTGTCAACCCGTTTACGTACCGGCA 355  
Db 1642 ACCCGCGCGGACACCGGCTGCGCAAGCTGTGCGCCGTTGATTCACATGCGCGG 1701  
Qy 356 TCGACTCTGCGCGCGCAAAATACAGCGACCGTGCACAGTGTGATGCTGCTCCG 415  
Db 1702 TCGATCTCTGCGCGCGCGGTCCAGGAGATGTGACGCGCTCGTGAAGCCATGTGG 1761  
Qy 416 GACAGAGAGTTCAGCTGTGCGGAGTTACCGGAGAGGATCCCGATGCGCGGATCA 475  
Db 1762 CGGCGCGCGGCGCGCGGATGTGATGAGATCTCTGCGCGCTGCGCGCGC--GATCA 1818  
Qy 476 GCGCTCTGTGAAAGTTCGCGCGGAGTGTGACGAGAAATTCCGTGCTTGGCTCGGCA 535

Db 1819 CCGTATCTCCGAATCTCTCGCGCTGCGCGAGCGGACCGCGCGCTTCCGCGCTCGGA 1878  
Qy 536 CTGCGCGCGCGCTGCGCGTGTGTTGATGCCCAAGTGTGATGAGAGACCAAGACCTCG 595  
Db 1879 CCGAGCGCTTGTCTTCCGACGATCCCGCGAGGCGCCAGACCG----- 1923  
Qy 596 TCGGCTCCGTCACGAGAGGGCTCGCGCTGTCTCATGACGTCTCTGATGAGCGCGCAGGA 655  
Db 1924 -----CATGGCCGAGATGAGCGGCTATCTCTCCGCGCTATGACCTCAAGCGCGG 1977  
Qy 656 ACCGCTCGAAATGACCTTTGACGATGTGCTTCAGGCGGAGCGACGCGACAGCG 715  
Db 1978 AGACGCGGAGAGACTGCTGACGCGCTGTGCGGACCGAGACGAGAGACGCTCCGCG 2037  
Qy 716 TGAGACGAGAGAGCTGTGCGCGCTCGGAGTATGCTGCTGCGCAACCGATACA 775  
Db 2038 TGACCTCGAGAGAGTGTGCTGTATGAGTCCCAATCTGCTGTGCGGAGCAGAGACCA 2097  
Qy 776 CGATCTACCTTATCGCGCTTCTGCTGTCAACTGTGCTGCGCGCGCGAGCGCTCGAGC 835  
Db 2098 CGGTAACTGATGTGCGCAACGCGCATGTACGCGCTGCTCGACCCGACAGCTGCGCG 2157  
Qy 836 TGTGAGAGCGGAGCGCGCTCATGAGAAAGCGCTCGATGAGTGTCTCGCTTCGACA 895  
Db 2158 CCTGCGGCGCGATGACGCTTTCGACGCGCGCGGTGAGAGATGTTGCGCTACGAG 2217  
Qy 896 ATATCTCAGAAATAGAACTGTGCTTTCGCCAGGACGAGACTGTGAGTACTGCGGGCAT 955  
Db 2218 GCCCGGTGAATTCGCGACCTACCGCTTCCGGGTGAGGCGGTGACCTGAGCGGACGG 2277  
Qy 956 CGATCAAGAAAGGAGATGCTTCTTCTGATCCCGAGCGCCTGAGATGAGGACTG 1015  
Db 2278 TCATCTCCGCGGTGACAGGATCTCTGCTGTCTGTGCGAGCGCCACCGCACCCCGGAGC 2337  
Qy 1016 TATTTCCAGGCGCAGCGTGTGATGTGCGACGGGACACGGGCGCGAGCCTCGGTACG 1075  
Db 2338 GCTTCCCGAGACCGGACCGCTTTCGACATCCGCGGAGACCGCGCGGACATCTCGCTTCG 2397  
Qy 1076 GTAGAGCCCCCATCTGCTGCCCCGGGAGTGTCTTCCTGCTCGAGCGGAGATGCGCG 1135  
Db 2398 GCCACGGGATCCACTTTCGATCTGCGCGCCCTTGGCGGTTGAGGCGCGGATCGCGC 2457  
Qy 1136 TGGGACCATCTTCTCGTGTGCTTCCCGAGATGAGTGAAGAACTCCCG 1186  
Db 2458 TCCGCGCCTTCTCGAACGCTGCGGACCTGCGCTGAGACGTCTTCCCGC 2508

## RESULT 15

US-09-657-440-21  
Sequence 21, Application US/09657440  
Patent No. 6509455  
GENERAL INFORMATION:  
APPLICANT: ASHLEY, Gary  
APPLICANT: BETLACH, Mary C.  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 300622002120  
CURRENT APPLICATION NUMBER: US/09/657,440  
CURRENT FILING DATE: 2000-09-07  
PRIOR APPLICATION NUMBER: 09/320,878  
PRIOR FILING DATE: 1999-05-27  
PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 21  
LENGTH: 5970  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-657-440-21



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:29:26 ; Search time 308.318 Seconds  
(without alignments)  
11031.749 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_62369\_63628

Perfect score: 1 atgcacagagagcagcga.....agccctccaaagctcgatag 1260

Sequence: 1 atgcacagagagcagcga.....agccctccaaagctcgatag 1260

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1258.4	99.9	68750	21	AAZ55887
2	1246.2	98.9	71989	21	AAA29349
3	149.2	11.8	67459	25	ABQ77491
4	143.8	11.4	1257	22	AAAF8157
5	131.6	10.4	1400	12	AAO14548
6	128.8	10.2	1347	25	ABX56047
7	123	9.8	5970	21	AAA75633
8	123	9.8	5970	21	AAZ56003

9	123	9.8	5970	25	ABZ56092
10	122.6	9.7	1191	24	ABN88910
11	121.4	9.6	1251	21	AAZ87301
12	120.8	9.6	1992	25	ABZ76672
13	119.6	9.5	36538	25	ABV75558
14	118.6	9.4	1449	24	ABN88911
15	116.8	9.3	1209	25	ABX56043
16	115	9.1	1879	12	AAQ11126
17	114.8	9.1	53789	19	AAV21187
18	114.6	9.1	1428	25	AAZ53024
19	114.4	9.1	1227	22	AAZ14511
20	114.4	9.1	1233	21	AAZ5790
21	114.4	9.1	12249	21	AAZ55840
22	114.4	9.1	15079	22	AAZ14499
23	114.4	9.1	18331	21	AAZ55857
24	112.8	9.0	15079	16	AAO91580
25	111.8	8.9	1293	25	AAZ53020
26	111.8	8.9	1293	25	AAZ53026
27	111.4	8.8	1428	25	AAZ53027
28	111.4	8.8	1700	14	AAQ45569
29	110	8.7	966	24	ABK74899
30	109.6	8.7	1224	21	AAZ5788
31	109.4	8.7	103599	23	ABX04971
32	108.6	8.6	1293	25	AAZ53031
33	107.8	8.6	135638	25	ABX34289
34	107	8.5	1293	25	AAZ53022
35	107	8.5	2795	22	AAZ17296
36	107	8.5	2795	22	AAZ1954
37	106.6	8.5	1413	25	AAZ53019
38	106.4	8.4	2918	21	AAZ53922
39	106.4	8.4	2919	21	AAZ54486
40	98	7.8	1293	25	AAZ53052
41	97.4	7.7	50937	21	AAZ09469
42	95.4	7.6	1293	25	AAZ53028
43	94.8	7.5	1998	12	AAQ11127
44	94.6	7.5	47981	22	AAZ50757
45	94	7.5	1293	25	AAZ53032

## ALIGNMENTS

AAZ55887	standard; DNA; 68750 BP.
AAZ55887	
10-APR-2000	(first entry)
Sorangium cellulosum 68.75 kb contig.	
Epithelone biosynthesis; type I polyketide synthase; taxol substructure; anticancer; ds.	
Sorangium cellulosum.	
Location/Qualifiers	
1..1826	
/*tag= a	
/product= "partial	
/note= "No initiation codon given in the specification"	
complement (1900..3171)	
/*tag= b	
/product= "Orf 2 protein (AAZ58581)"	
3415..5556	
/*tag= c	
/product= "Orf 3 protein (AAZ58582)"	
complement (5612..5592)	
/*tag= d	
/product= "Orf 4 protein (AAZ58583)"	
6226..6675	

Contig 002 from co  
Streptomyces sp. c  
S. venezuelae macr  
Streptomyces sp. T  
Saccharopolyspora  
Streptomyces sp. c  
M. echinospora cal  
Sequence encoding  
Amycolatopsis medl  
Streptomyces albol  
Streptomyces clav  
Hydroxylase orf3 1  
Complete Mitomycin  
Streptomyces clav  
Complete nucleotid  
S. clavuligerus cl  
Streptomyces lydic  
Streptomyces kaug  
Streptomyces rimos  
Sequence soyC and  
Bacillus lichenifo  
Hydroxylase orf4 1  
S. cinamomensis m  
Streptomyces lydic  
S. atroolivaceus 1  
Streptomyces chat  
Streptomyces grise  
Streptomyces grise  
Streptomyces rimos  
S. tendae nikomyc  
Streptomyces tende  
Streptomyces tuber  
Streptococcus olea  
Streptomyces tuber  
Sequence encoding  
Micromonospora meg  
Streptomyces lydic

```

FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
CDS     7610..11875
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
CDS     11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
CDS     16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
CDS     21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
CDS     43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
CDS     54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
CDS     62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
CDS     63779..66333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
CDS     complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
CDS     64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
CDS     complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
CDS     65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
CDS     complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
CDS     complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
CDS     66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
CDS     67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
CDS     68346..68750
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX
XX      WO996028-A2.
XX
XX      23-DEC-1999.
XX
XX      16-JUN-1999.
XX      99WO-EP04171.
XX
XX      18-JUN-1998.
XX      98US-0099504.
XX      24-SEP-1998.
XX      98US-0101631.
XX      05-FEB-1999.
XX      99US-0118906.

```

```

PA      (NOVS ) NOVARTIS AG.
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX
XX      MPI: 2000-097741/08.
DR      P-PSDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
DR      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
DR      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
DR      AAV58592, AAV58593, AAV58594.
XX
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy -
XX
XX      Claim 14; Page 87-104; 174pp; English.
XX
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-78) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiply drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
XX
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX      Query Match          99.9%; Score 1258.4; DB 21; Length 68750;
XX      Best Local Similarity 99.9%; Pred. No. 3.1e-279;
XX      Matches 1259; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX      1  ATGACACAGGAGCAAGCCGATCAGATGAGAGCAAGCCGCTTGATTCGAAGCCGTTG 60
DB      62369  ATGACACAGGAGCAAGCCGATCAGATGAGAGCAAGCCGCTTGATTCGAAGCCGTTG 62428
XX
XX      61  GCGCCTGGGTAAGCGGAGAGACCCGTTCCCGGATGAGCGGCTGAGAGGCAACCCCC 120
DB      62429  GCGCCTGGGTAAGCGGAGAGACCCGTTCCCGGATGAGCGGCTGAGAGGCAACCCCC 62488
XX
XX      121 ATCTTCTACTGGAGTGAAGCGCGCTCTCGGCTCTACCCGATACAGAGCTGCGGCG 180
DB      62489  ATCTTCTACTGGAGTGAAGCGCGCTCTCGGCTCTACCCGATACAGAGCTGCGGCG 62548
XX
XX      181 GGTTCGCGAGCAAGCGCTTCGCGTCAAGTGAAGAGTGGGAATTCGAGCGGAGTAC 240
DB      62549  GGTTCGCGAGCAAGCGCTTCGCGTCAAGTGAAGAGTGGGAATTCGAGCGGAGTAC 62608
XX
XX      241 TCGTCGGCCATTCGCGAGCTCAGCGATATGAAGAGTACGATTTGTCGGGCTGCGCGG 300
DB      62609  TCGTCGGCCATTCGCGAGCTCAGCGATATGAAGAGTACGATTTGTCGGGCTGCGCGG 62668
XX
XX      301 GAGGATCAGCGCTCGGCTCGGCAAGCTGTCACACCCGTCGTTACCTACGCGCCATCAG 360
DB      62669  GAGGATCAGCGCTCGGCTCGGCAAGCTGTCACACCCGTCGTTACCTACGCGCCATCAG 62728
XX
XX      361 CTGCTGGCGGCGGAATACAGGCAAGCTGACCAAGCTGCTGATGCTGCTCGGACAA 420

```

Db	62129	CTGCTGCGCGCCGAAATACACGCCACCGTGCACCAAGTCTGCTGATCTCGGACAA	627888
Qy	421	GAGAGATTGCAAGTTGTGTGCGGATTAACCGGAGGAATCCGATGCGCGATCAAGCT	480
Db	62789	GAGAGATTGCAAGTTGTGTGCGGATTAACCGGAGGAATCCGATGCGCGATCAAGCT	628488
Qy	481	CTGTTGAAGGTTCCGGCCGACTGTGACGAAAGTTCCGTGCTTCGGCTCGCGACTGGC	540
Db	62849	CTGTTGAAGGTTCCGGCCGACTGTGACGAAAGTTCCGTGCTTCGGCTCGCGACTGGC	629088
Qy	541	CGCCGCTCGGCGGAGTTGTGTGTCGCCAGTGCATGAGGAGCAAGACCCGCGCGCG	600
Db	62909	CGCCGCTCGGCGGAGTTGTGTGTCGCCAGTGCATGAGGAGCAAGACCCGCGCGCG	629688
Qy	601	TCCGTCAACGAGGGGCTCGCGCTGCTCATGACCTCTCGATGAGCGCGCAGAAACCG	660
Db	62969	TCCGTCAACGAGGGGCTCGCGCTGCTCATGACCTCTCGATGAGCGCGCAGAAACCG	630288
Qy	661	CTCGAANAATGACGCTTGTGACGATGTGCTTCAGGCGGACGACGCGACAGGCTGAGC	720
Db	63029	CTCGAANAATGACGCTTGTGACGATGTGCTTCAGGCGGACGACGCGACAGGCTGAGC	630888
Qy	721	ACGAGAAGAGCTGTGCGGCTCGTGTGGGTGCATTAATGCTGCTGAGCACCGATACAGATC	780
Db	63089	ACGAGAAGAGCTGTGCGGCTCGTGTGGGTGCATTAATGCTGCTGAGCACCGATACAGATC	631488
Qy	781	TACCTTAATGCGGTTGCGTGTGCTCAACCTGCTCGGTCGCGCCGAGGCGCTCGAGCTGTG	840
Db	63149	TACCTTAATGCGGTTGCGTGTGCTCAACCTGCTCGGTCGCGCCGAGGCGCTCGAGCTGTG	632088
Qy	841	AAGGCCAGACCCCGGACTCATAGAGAAACGCGCTCGATGAGGTGCTCCGCTTCGCAATATC	900
Db	63209	AAGGCCAGACCCCGGACTCATAGAGAAACGCGCTCGATGAGGTGCTCCGCTTCGCAATATC	632688
Qy	901	CTCGAATAGGAATCTGTGCGTTTCGCGCAGGACAGACCTGGAGTAAGTCGCGGAGATGATC	960
Db	63269	CTCGAATAGGAATCTGTGCGTTTCGCGCAGGACAGACCTGGAGTAAGTCGCGGAGATGATC	633288
Qy	961	AAGAAAGGGAGATGATCTTTCTTCCTGATCCGAGCGCCCTGAGAGATGGGACTGATTC	1020
Db	63329	AAGAAAGGGAGATGATCTTTCTTCCTGATCCGAGCGCCCTGAGAGATGGGACTGATTC	633888
Qy	1021	TTCAGGCGCAACGTTGTGATGTGTGCACGGGACACGGGCGCGAGGCTCCGCTACGGTATA	1080
Db	63389	TTCAGGCGCAACGTTGTGATGTGTGTGCACGGGACACGGGCGCGAGGCTCCGCTACGGTATA	634488
Qy	1081	GGCCCCCATGTCTGCCCGGGGTGTCCCTTGCTCGCTCGAGGCGGAGATCGCCGTGGCC	1140
Db	63449	GGCCCCCATGTCTGCCCGGGGTGTCCCTTGCTCGCTCGAGGCGGAGATCGCCGTGGCC	635088
Qy	1141	ACCATCTTCGTAGGTTCCCGAGATGAAAGCTGAAAGAACTCCGCTGTTTGATATCAAC	1200
Db	63509	ACCATCTTCGTAGGTTCCCGAGATGAAAGCTGAAAGAACTCCGCTGTTTGATATCAAC	635688
Qy	1201	CCCGCGTTCCGGAAACATCGAATCACTTAAGCTCATCTTGAAGCCCTTCAAAGTGTGATG	1260
Db	63569	CCCGCGTTCCGGAAACATCGAATCACTTAAGCTCATCTTGAAGCCCTTCAAAGTGTGATG	636288

RESULT 2  
AAA29349  
ID AAA29349 standard; DNA, 71989 BP.  
XX  
AC  
XX AAA29349;  
DT 12-SEP-2000 (first entry)  
XX  
XX Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.  
DE  
XX  
XX  
KW Epothilones; polyketide synthase; epoA; epoB; epoc; epod; epoF; epof;  
KW epouI; epok; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal  
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.

XX	OS	Sorangium cellulosum.
XX	Key	Location/Qualifiers
FH	3..992	
FT	CDS	
FT		/tag= a
FT		/label= ORF_A
FT		/product= transposase
FT		/notes= "not part of the PKS"
FT		989..1501
FT	CDS	
FT		/tag= b
FT		/label= ORF_B
FT		/product= transposase
FT		/note= "not part of the PKS"
FT		1998..6263
FT	CDS	
FT		/tag= c
FT		/label= epa gene
FT		/note= "encodes the loading domain"
FT		2031..3548
FT	misc_RNA	
FT		/tag= d
FT		/note= "encodes ketide synthase (KS-Y) of the loading domain"
FT		3621..4661
FT	misc_RNA	
FT		/tag= e
FT		/note= "encodes acyl transferase (AT) of the loading domain"
FT		4917..5810
FT	misc_RNA	
FT		/tag= f
FT		/note= "encodes enoyl reductase (ER) of the loading domain, potentially involved in formation of the thiazole moiety"
FT		5856..6155
FT	misc_RNA	
FT		/tag= g
FT		/note= "encodes acyl carrier protein (ACP) of the loading domain"
FT		6260..10493
FT	CDS	
FT		/tag= h
FT		/label= epob_gene
FT		/note= "encodes module 1, the NRPS module"
FT		2031..3548
FT	misc_RNA	
FT		/tag= i
FT		/note= "encodes condensation domain C2 of the NRPS module"
FT		2031..3548
FT	misc_RNA	
FT		/tag= j
FT		/note= "encodes condensation domain C2 of the NRPS module"
FT		6861..6887
FT	misc_RNA	
FT		/tag= k
FT		/note= "encodes heterocyclization signature sequence"
FT		6861..6887
FT	misc_RNA	
FT		/tag= l
FT		/note= "encodes condensation domain C4 of the NRPS module"
FT		7358..7366
FT	misc_RNA	
FT		/tag= m
FT		/note= "encodes condensation domain C7 (partial) of the NRPS module"
FT		7898..7921
FT	misc_RNA	
FT		/tag= n
FT		/note= "encodes adenylation domain A1 of the NRPS module"
FT		7898..7921
FT	misc_RNA	
FT		/tag= o
FT		/note= "encodes adenylation domain A1 of the NRPS module"
FT		8261..8308
FT	misc_RNA	
FT		/tag= p
FT		/note= "encodes adenylation domain A3 of the NRPS module"
FT		8411..8422
FT	misc_RNA	
FT		/tag= q
FT		/note= "encodes adenylation domain A4 of the NRPS module"
FT		8861..8905
FT	misc_RNA	
FT		/tag= r
FT		/note= "encodes adenylation domain A6 of the NRPS module"



```

FT misc_RNA 8966..8983
FT /tag= s
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS module"
FT CDS 10639..16137
FT /tag= x
FT /label= epoc_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epod_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574

```

```

FT /tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287
FT /tag= ap
FT /note= "encodes ER5"
FT misc_RNA 30321..30869
FT /tag= aq
FT /note= "encodes KR5"
FT misc_RNA 31077..31373
FT /tag= ar
FT /note= "encodes ACP5"
FT misc_RNA 31440..32807
FT /tag= as
FT /note= "encodes KS6"
FT misc_RNA 33018..34067
FT /tag= at
FT /note= "encodes AT6"
FT misc_RNA 34107..34676
FT /tag= au
FT /note= "encodes DH6"
FT misc_RNA 35760..36641
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA 36705..37256
FT /tag= aw
FT /note= "encodes KR6"
FT misc_RNA 37470..37769
FT /tag= ax
FT /note= "encodes ACP6"
FT CDS 37912..49308
FT /tag= ay
FT /label= epof_gene
FT /note= "encodes modules 7 and 8"
FT misc_RNA 38014..39375
FT /tag= az
FT /note= "encodes KS7"
FT misc_RNA 39589..40626
FT /tag= ba
FT /note= "encodes AT7"
FT misc_RNA 41341..41922
FT /tag= bb
FT /note= "encodes KR7"
FT misc_RNA 42181..42423

Query Match 98.9%; Score 1246.2; DB 21; Length 71989;
Best Local Similarity 99.4%; Pred. No. 2e-276;
Matches 1251; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGACACAGAGCAGCAGATCAGATGAGAGCAGAGCCTGCTTTCGACTTCAAGCCGTTTC 60
DB 56757 ATGACACAGAGCAGCAGATCAGATGAGAGCAGAGCCTGCTTTCGACTTCAAGCCGTTTC 56816
QY 61 GCGCCTGGGTACGCGAGAGACCCGTTCCCGCGATCGAGCGCTGAGAGGCAACCCCC 120
DB 56817 GCGCCTGGGTACGCGAGAGACCCGTTCCCGCGATCGAGCGCTGAGAGGCAACCCCC 56876
QY 121 ATCTTCTACTGGAGTAGAAGCCGCTCCGAGGCTCCACCCGATACACAGAGTGTGCGCG 180
DB 56877 ATCTTCTACTGGAGTAGAAGCCGCTCCGAGGCTCCACCCGATACACAGAGTGTGCGCG 56936
QY 181 GTGTCGCGAGCAGAGCTTCGCGGTAGTCAGAGAGAGTGGGAATCGAGCCGGAAGTAC 240
DB 56937 GTGTCGCGAGCAGAGCTTCGCGGTAGTCAGAGAGAGTGGGAATCGAGCCGGAAGTAC 56996
QY 241 TCGTCGCGCATTCGCGAGCTCAGCGATATGAAGAGTACGAGTTGTGCGGCTGCGCGCG 300
DB 56997 TCGTCGCGCATTCGCGAGCTCAGCGATATGAAGAGTACGAGTTGTGCGGCTGCGCGCG 57056
QY 301 GAGATCACGCTCGGCTCGCAAGCTCGTCAACCCGCTGTTTACGTCAAGCCCATTCAGC 360
DB 57057 GAGATCACGCTCGGCTCGCAAGCTCGTCAACCCATGTTTACGTCAAGCCCATTCAGC 57116
QY 361 CTGCTGCGCGCCGGAATACAGCGCACCGTCAAGCAGTGTGATGATGCTGCTCCGAGCA 420

```

```

Db 57117 CTGCGCCGCGCGAATACAGGCGACCGTCGACGCTGCTGATGCTCGCTCCGAGCAA 57176
Oy 421 GAGGAGTTGACGATTGTCGGGATTAACGGGAGGGAATCCCGATCCGGGCGATCGAGCCT 480
Db 57177 GAGGAGTTGACGATTGTCGGGATTAACGGGAGGGAATCCCGATCCGGGCGATCGAGCCT 57236
Oy 481 CTGTTGAAGGTTCCGGGCGAGTGTGACGAGAAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 540
Db 57237 CTGTTGAAGGTTCCGGGCGAGTGTGACGAGAAAGTTCCGTCGCTTCGGCTCGGCGACTGCG 57296
Oy 541 CGCGGCTCGCGCTGGGTTTGGTGTCCCAAGTTCATGAGAGACCAAGACCTGTGTCCG 600
Db 57297 CGCGGCTCGCGCTGGGTTTGGTGTCCCGGTCGATGAGAGACCAAGACCTGTGTCCG 57356
Oy 601 TCCGTCACCGAGGGGCTGGCGCTGCTCATGACGTCCTCGATGAGCGGGCGAGAGACCCG 660
Db 57357 TCCGTCACCGAGGGGCTGGCGCTGCTCATGAGCGTCCTCGATGAGCGGGCGAGAGACCCG 57416
Oy 661 CTCGAAATGACGTCCTGACGATGCTGCTTCAAGCGCGGCGAGCGAGAGGCTGAGC 720
Db 57417 CTCGAAATGACGTCCTGACGATGCTGCTTCAAGCGCGGCGAGCGAGAGGCTGAGC 57476
Oy 721 ACGAAGAGCTGTGCTGCGCTCGTGGGTGCGATTATGCTGCTGCGACCGATACAGATC 780
Db 57477 ACGAAGAGCTGTGCTGCGCTCGTGGGTGCGATTATGCTGCTGCGACCGATACAGATC 57536
Oy 781 TACCTTATCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 57537 TACCTTATCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57596
Oy 841 AAGGCGGAGCGCGGCTCATGAGGAAACCGCTCATGAGTGTGCTGCTGCTGCTGCTGCTGCTG 900
Db 57597 AAGGCGGAGCGCGGCTCATGAGGAAACCGCTCATGAGTGTGCTGCTGCTGCTGCTGCTGCTG 57656
Oy 901 CTCGATATGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 57657 CTCGATATGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57716
Oy 961 AAGAAAGGAGGATGCTTCTCTGATCCCGAGCGGCTGAGAGATGGAATGCTATTC 1020
Db 57717 AAGAAAGGAGGATGCTTCTCTGATCCCGAGCGGCTGAGAGATGGAATGCTATTC 57776
Oy 1021 TCCAGGCGAGAGCTGTTGATGTCGAGCGGACACGCGGCGCGAGCTGCGTATGAG 1080
Db 57777 TCCAGGCGAGAGCTGTTGATGTCGAGCGGACACGCGGCGCGAGCTGCGTATGAG 57836
Oy 1081 GGGCCCCATGTCGCCCCGGGGGTGTCCTTGTGCTGCTGAGCGGAGATGCGCGTGAGC 1140
Db 57837 GGGCCCCATGTCGCCCCGGGGGTGTCCTTGTGCTGCTGAGCGGAGATGCGCGTGAGC 57896
Oy 1141 ACCATCTTCCTAGTTCCTCCGAGATGAGAGTGAAGAAATCCCGTGTGTTGATACAC 1200
Db 57897 ACCATCTTCCTAGTTCCTCCGAGATGAGAGTGAAGAAATCCCGTGTGTTGATACAC 57956
Oy 1201 CCCGCGTCCGGAACATGCAATCAACGTCATCTTGAAGCCCTCCAAAGCTGAGATA 1259
Db 57957 CCCGCGTCCGGAACATGCAATCAACGTCATCTTGAAGCCCTCCAAAGCTGAGATA 58015

```

RESULT 3  
ABQ77491/c  
ID ABQ77491 standard; DNA; 67459 BP.

XX ABQ77491;

XX 14-MAY-2003 (first entry)

DE S. aurentiaca DNA containing sti gene cluster.

KM stiA; stiB; stiC; stiD; stiE; stiF; stiG; stiH; stiJ; stiK; stiL;  
KM aromatic structure biosynthesis; beta-ketoacyl synthase;  
KM acyltransferase; dehydratase; enoylreductase; beta-ketoacylreductase;

```

KW acyl carrier protein; O-methyl transferase; de.
XX Stigmatella aurentiaca.
XX OS
XX Key Location/Qualifiers
XX CDS 14345..21466
XX FT /tag= a
XX FT /product= "stiA"
XX FT /note= "GTG start codon"
XX FT 21463..26220
XX FT /tag= b
XX FT /product= "stiB"
XX FT 26284..31941
XX FT /tag= c
XX FT /product= "stiC"
XX FT 31934..37738
XX FT /tag= d
XX FT /product= "stiD"
XX FT 37740..43553
XX FT /tag= e
XX FT /product= "stiE"
XX FT 43546..50202
XX FT /tag= f
XX FT /product= "stiF"
XX FT 50205..54401
XX FT /tag= g
XX FT /product= "stiG"
XX FT 54398..59212
XX FT /tag= h
XX FT /product= "stiH"
XX FT 59223..63002
XX FT /tag= i
XX FT /product= "stiJ"
XX FT 63070..63840
XX FT /tag= j
XX FT /product= "stiK"
XX FT 63929..65470
XX FT /tag= k
XX FT /product= "stiL"
XX FT /note= "GTG start codon"
XX DE10128661-A1.
XX 19-DEC-2002.
XX PD
XX 15-JUN-2001; 2001DE-1028661.
XX PE
XX 15-JUN-2001; 2001DE-1028661.
XX PR
XX 15-JUN-2001; 2001DE-1028661.
XX PA
XX (GBFB ) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX PI Kunze B, Nordstiek G, Sliakowski B, Bioecker H, Galtatzis N;
XX Muehler R;
XX DR WPI; 2003-185404/19.
XX XX
XX Nucleic acid molecule comprising one or more genes from Stigmatella
XX species -
XX PS
XX Claim 3; Page 9-43; 48pp; German.
XX CC This invention describes a novel nucleic acid molecule comprising one or
XX CC more genes from Stigmatella species selected from stiA, stiB, stiC, stiD,
XX CC stiE, stiF, stiG, stiH, stiJ, stiK and stiL. The nucleic acid is used for
XX CC biosynthetic purposes, particularly for producing aromatic structures.
XX CC The products of the invention have domains sharing high homology with
XX CC beta-ketoacyl synthase (KS), acyltransferase (AT), dehydratase (DH),
XX CC enoylreductase (ER), beta-ketoacylreductase (KR), acyl carrier protein
XX CC (ACP), O-methyl transferase (O-MT). This sequence represents a
XX CC polynucleotide isolated from S. aurentiaca which contains the sti gene
XX CC cluster described in the disclosure of the invention.
XX Sequence 67459 BP; 10755 A; 20921 C; 23348 G; 12435 T; 0 other;

```

Query Match	11.8%	Score 149.2	DB 25	Length 67459;
Best Local Similarity	49.2%	Pred. No. 1.7e-24;		
Matches 450;	Conservative 0;	Mismatches 456;	Indels 6;	Gaps 2;
QY	298	CCGGAGATCA	CAGCTCGGGTCCGACAGCTCGTCAACCCGCTTTACGTACCGCCATC	357
Db	5928	CCGCGCTGCACACGCGGGTGC	CGCACTGATCGTGGGCGCGCTGACGCCCGCGCTTTG	5869
QY	358	GACCTGCTGCGCGCCGAAAT	TACCGCACCGTCCACCAAGCTGCTCGAATGCTCGCTCCGGA	417
Db	5868	TCGACAAATGAGCAGAGGAT	TCGCAACCTGTGACCGCTGTGAGAGGGCTGGCTGTG	5809
QY	418	CAGAAGAGTTGACAGTTGT	TCGGGATTAACCGAGAGGAATCCCGATCCGCGCATCAGC	477
Db	5808	AAGGGCCCGGACAGACTGAT	TCAGAGACTTCGCGCCCGCATCCCATGAGATTAATGGC	5749
QY	478	GCTCTGTGAAGTTCCGGC	CCGAGTGTGACAGAAATTCGCTCGCTCCGCGACT	537
Db	5748	AACCTGTGACATCCCGAC	CGGAGAGCGGGGCGCTGCGGGCTGTGCTCGCGCATC	5689
QY	538	GCGGCGCGCTTCGGCGT	TGGGTTTGGTGCCTCCAGGTCCATGAGAGACCAACCTGGTC	597
Db	5688	CTGCGCGCGCTCGAAC	CCCCGTGACCGCCAGAGAGAGCGCTCGGAAATGAGCGGTA	5629
QY	598	GCTGCCGTCACCGAGGGG	CTCGCGCTTCATGACGCTTCGATGACGCGCAGAAC	657
Db	5628	ACAGAGTTCTCGAGTAC	TACTCCGATCTCTGTCGCGACCGCGACCGCGCCCGGGGAC	5565
QY	658	CCGCTCGAAATGACGT	TTGACATGTGCTTACGCGAGGCGCCAGCGACAGAGCTG	717
Db	5568	CC---CGCAACGAGCGT	GTCTACGCGCTTACAGGGCGAATCGAGGGGAAAGATTTG	5512
QY	718	AGCAGGAAGAGCTG	GTGCGCTGTGGGTGCATTAATCGCTGCGACCGATACAG	777
Db	5511	ACCGAGACGAGCTG	CTGTCACACAGTGCCTTTCTCTGAACGGGGGACAGAGACGA	5452
QY	778	ATTACCTTAATCGCG	TTGCTGTGCTCAACCTGTGCGCGGTGACCCCGAGCGCTCGAGCTG	837
Db	5451	ACGAACCTCATCGG	GAACGCGCTGGAACTGTGCGGCGCTTCCGGACGAGCGCGCTCGG	5392
QY	838	GTEAAGCCGAGCC	CCGGGCTATGAGAGAACCGCTTCATGAGAGTCTCCGCTTGAACAAT	897
Db	5391	CTGTGTGAGCGCC	CCCGCGCTGATCCCGACGGCAGTGAAGAGTCTGTGCTTAAGAAATCC	5332
QY	898	ATCTCTCAAAATAG	GAATCTGTGCTGCGCAGCGACGACCTGAGATCATGCGGGGCATCG	957
Db	5331	TTCAACCAAGCT	TCGGCACCGCGGGGTGC---GAGAGACACGAGATTTGGCGGTGACG	5275
QY	958	ATCAAGAAAGGG	AGATGATCTTTCTCTCGTATCCGAGCGGCTGTGAGATGGGACTGT	1017
Db	5274	GTCGCGACCGG	CACCTTCTCAGCCTTGACATCGGCGCGCCGACCGGACCCCGCGGC	5215
QY	1018	TTCTTCACGAGCC	AGAGTGTGATGTGAGACGGGACAGGGGCGGAGGCTTGCGTACGCT	1077
Db	5214	TTTCAGAGACCC	AGACACTTGACGTGGGGGCGCACGCCCAACCGGCACTTGCTTCGCG	5155
QY	1078	AGAGGCCCCCA	TGATCTGTGCGCCGGGGTGTCCCTTGTGCTGCTCGAGAGCGAGATCGCCGTG	1137
Db	5154	GCGGGGGCGCAT	AGCTGTGCGCGGGATGAACCTTGTGCGCGAATGAGAGCCCGGATACGTTG	5095
QY	1138	GGACACATCTT	TCGTAAGTTCCCGAGATGAAAGCTGAAGAAACTCCGCTGTTGGATAC	1197
Db	5094	GCGGCGTTTCT	TCGCGCGCGCTTCCCGGACTATGCGGTACCGGACCGCGCGGTGAGGCGCGG	5035
QY	1198	CACCCCGG	GTCCG 1211	
Db	5034	CGGGGCGG	TTCCG 5021	

RESULT 4  
AAF81357  
ID AAF81357 standard; DNA; 1257 BP.

XX	AA81357;
AC	
XX	
DT	04-JUN-2001 (first entry)
XX	
DE	Quorum sensing controlled gene gacI22 ORF.
XX	
XX	Quorum sensing; antibacterial; bacterial signalling;
KW	opportunistic pathogen; immunocompromised; burn; cystic fibrosis;
KW	immunosuppressive therapy; AIDS; ss.
XX	
OS	Pseudomonas aeruginosa.
XX	
XX	MO200118248-A2.
PD	
XX	15-MAR-2001.
PF	
XX	01-SEP-2000; 2000MO-US24141.
PR	
XX	03-SEP-1999; 99US-0153022.
PA	
XX	(IOWA ) UNIV IOWA RES FOUND.
XX	(QUOR-) QUORUM SCI INC.
PI	
XX	Whiteley M, Lee XM, Greenberg EP, Muh U;
DR	
XX	WPI; 2001-265973/27.
PT	
XX	Identifying modulators of quorum sensing signaling in Pseudomonas
PT	aeruginosa bacteria, useful for treating infections in
PT	immunocompromized patients -
PS	
XX	Claim 46; Page 101-102; 115pp; English.
XX	
CC	Bacteria signal to one another to coordinate expression of specific genes
CC	in a cell density dependent fashion". This "bacterial signalling" is
CC	called "quorum sensing and response". Quorum sensing allows a bacterial
CC	species to sense its own number and regulate gene expression according to
CC	population density. The present sequence is an open reading frame (ORF)
CC	of a Pseudomonas aeruginosa quorum sensing controlled gene. Inhibitors of
CC	quorum sensing signaling renders a bacterial population more susceptible
CC	to treatment. The present invention relates to a method for identifying
CC	modulators of quorum sensing signaling in Pseudomonas aeruginosa
CC	bacteria. Modulators of quorum sensing may be used to treat P.
CC	aeruginosa infections. P. aeruginosa is an opportunistic pathogen of
CC	immunocompromised individuals (burn patients, cystic fibrosis patients,
CC	patients undergoing immunosuppressive therapy and patients with AIDS).
CC	
SO	Sequence 1257 BP; 210 A; 469 C; 402 G; 176 T; 0 other;
Query Match	11.4%; Score 143.8; DB 22; Length 1257;
Best Local Similarity	48.7%; Pred. No.1.5e-23;
Matches 455; Conservative	0; Mismatches 447; Indels 33; Gaps 1
OY	254 CCGAGCTCAGCGAATGAAAGTAGTCGATTGTTGGAGCTGCCGCCGAGATCAGCTC 313
DB	233 CGGGAGTGGCGAGGAGACTGACCAACCAATGCTCACTCGAACCCGCCGACCAATACCC 292
OY	314 GGGTCCGGAAGCTGCTCAACCCGTCGTTAAGTACCGGCCATCGACTGACTGTCGCCGCCG 373
DB	293 GCTGCGCTCGCTGGTGGCCGCCGGTTCACCCCGCCGACAGTGAGCGCTTCAACCCG 352
OY	374 AATACAGCGACCGTGCAGCAGCTGCTCGATGCTCGCTCCGGACAAGAGAGTTGACG 433
DB	353 ATATGACGATCACCGAGGCAATGCTGAGCGCATGCGGCCGCCGAAACAGGCCGAC 412
OY	434 TTGTGCGGATTAACCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTGAAGTTTC 493
DB	413 TGATGCGCGAATTCGCGCGATCCCGCTGACCAATCGCGGATATCTTCAGAGTGTGCGGCAATTC 472
OY	494 CGGCGAATGTGACGAGAAAGTTTCGCTCGCTTCGCTCGGCGACATCGCGCGGCTCGCG 553
DB	473 CCGAGCGCGAGCGGCAACACGCCGCCCAAGTCTG----- 506

QY	554	TGGGTTTGATGATCCCGCAGGCTGCATGTAGAGAGAACAAACAGCTGGATTCGGCTCCGTCACAGAG	613
Db	507	-----GGAGCGCCAGGCGGAACTGTCTGTGTGCGCGGAGAGAGGCCACAGGCGCTTGGCGCGATG	559
QY	614	GAGTCGCGCTGCTGCATGACGTCCTTCGATGAGCGGGCGCGAGAAACCCGCTTCGAAATGACG	673
Db	560	CGCAGGTGCACTACTGGCGCGGTGCTGTGAGGGCCAAAGCGCGCGGACGCGCGCCGACGACG	619
QY	674	TCTTGACGATGCTGTGTTGAGGCGCGGACGCCGACGACAGGCTTGAGCAGCAAGAGGCTTG	733
Db	620	TCTACAGCGGGGTGTGTGACGAGCGCGCGACGAGAGGCGGCCAGTTGAGCGAAGCGAACTCG	679
QY	734	TGCGGCTGCTGGGTGCAGATTATCGCTGTGCGACCGATACACAGATCTTACTTTATCGCGT	793
Db	680	TCTCATATGGCCCACTGCTGATGATGAGGGCGTTGAGACACACATGAACTGATTCGGCA	739
QY	794	TGCGTGTGCTCAACTGCTGCGGTCGCCCGGAGCGCGCTGAGCTGTGTAAGGCGAGCCCG	853
Db	740	ACGCGCTGTGACCTCTGCTGTGTCAACCCGAGCAACTGCGCTTGTGCGCGGCGCAGCCGG	799
QY	854	GGCTCATAGGAGACGCGCTGCATGAGGTGTCGCGTTGAGCAATTATCCCAATAAGAA	913
Db	800	AACCTCTGCCAACGCCCAATGAGAACTGTGTCCGCAACACAGTCCGGTGCCTCGCA	859
QY	914	CTGTGCGTTTGCCGACGAGCAAGCACTGTGAGTACTGTGCGGGGCATGCATCAAGAAAGGAGAA	973
Db	860	TGTTGCGCTTCAACGATGAGAACGTGGAACGTGACGGGGTCACTATCCCCCGCGGAT	919
QY	974	TGTGTTTTCTCTGATTCGCCAGCGCCCTGTGAGAGATGGGACTATTTCTTCAGGCCACAGC	1033
Db	920	ACATTCCTGTGATCTTCCAACTGTGACCGGCAACACAGATGACCGAGCGCTTGCAGATTCGCCAC	979
QY	1034	TGTTTGATGTGTGACGAGGACACAGGCGCGGAGCGCTGCGGTAGACGATAGAGGCCCCCAATGTC	1092
Db	980	GCTTCGACTCACCGCGCAACCCGATGCGCATCTGTGCGCTTTCGCGCTTTCGCGCTGCACTACT	1033
QY	1094	GCCCCGGGGTGTCCCTTGTGCTGTGACCTGCAGAGCGGAGATGCGCGGTGGGACCAATCTTCCGTA	1155
Db	1040	GCGTCGGCGGCTCGCTGTGGCCGGGCTGTGAGGGGCGGATGCGCATTCACAGCGCTGTCTGCGC	1092
QY	1154	GCTTCCCGGAGTGAAGCTGAAGAAATCTCCGCTG	1188
Db	1100	GCTTCCCGGACTTCCAGTTGGGGGTGCCCCACAGCG	1134

	RESULT	5
XX	AAQ14548	
ID	AAQ14548 standard; DNA; 1400 BP.	
XX		
AC	AAQ14548;	
XX		
DT	28-JAN-1992 (first entry)	
DE	EryF gene.	
XX		
KW	C-6 hydroxylation; erythromycin; 6-deoxyerythromycin; antibiotics,	
RN	saccharopolyspora; cytochrome P450 monooxygenase; ss.	
XX		
OS	Saccaropolyspora erythraea.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	158..1372
FT		/+tag= a
XX		
PN	WO9116334-A.	
XX		
PD	31-OCT-1991.	
XX		
PF	16-APR-1991;	91WO-USO2600.
XX		
PR	18-APR-1990;	90US-0510483.
XX		

PA (ABBO ) ABBOTT LABORATORIES.  
XX  
PI Weber JM;  
XX  
DR WPJ; 1991-339744/46.  
DR P-PsDB; AAR14724.  
XX  
PT New 6-deoxy-erythromycin derivs. - are antibiotics with increased  
PT acid stability, produced by cultivation of saccharopolyspora.  
XX  
PS Disclosure; Fig 3; 56pp; English.

The *eryF* gene encodes the 6-hydroxylase component of the cytochrome P450 monooxygenase system responsible for the hydroxylation of 6-deoxyerythronolide B to erythronolide B. Interruption of this step results in the formation of deoxyerythromycin A and new derivatives useful as antibiotic which have better stability against acids than the corresponding erythromycins. Interruption of the reaction can be effected by an insertion into the *eryF* gene of a plasmid, gene replacement or chemical or light-induced mutagenesis.

The gene is positioned between the *eryH* and *eryG* genes.  
See also AAQ14549.

Sequence 1400 BP; 212 A; 495 C; 479 G; 214 T; 0 other;

Query Match	10.4%;	Score 131.6;	DB 12;	Length 1400;
Best Local Similarity	-49.1%;	Pred. No. 9.5e-21;		
Matches 430;	Conservative 0;	Mismatches 419;	Indels 27;	Gaps 2

OY 298 CCGGAGGATCAACGCTGGGATCCGGAAGGCTGTCAACCCGCTGTTATCTCAACGCCCACTC 357  
 Db 440 CCGCGAACCCACACCCGGCTGCGCAAGCTGGATGTGCGAGGATTCACCGTCCGCCGCGTG 499  
 OY 358 GACCTGCTGCGCGCGGAATATCAGCGCACCGTCCAGACAGTGTCTGATGCTTCGTCGGA 417  
 Db 500 GAGCGGATGCGGCCCCCGCTCGAGACAGATCAACCGGAGGCTGTCTCGAGAGGTGGCGAC 559  
 OY 418 CAAGAGGAGTTGAGCTTTGTGCGGGATTAAGCGCGAGGGGAATCCCGATCGCGCGATCAGC 477  
 Db 560 TCCGCGCGGTGCGACATGTCTTACCGCTTCCGCCACCGCGTGGCCATCAAGTCAATCTGC 619  
 OY 478 GCTCTGTTGAAGTTCCCGGCGGAGTGTGACGAGAAGTTCGCTCGCTTCGAGTCCGAGCACT 537  
 Db 620 GAGCTGCTCGGCGGTGACGAGAGAATACGGGGGGAGTTCCGGGCGGTGAGCTCCGGA-- 677  
 OY 538 GCGGCGCGGCTCGCGGTTGGTTGGTGTGCTCCAGGTGATGAGAGAGACCAAGACTGTGTC 597  
 Db 678 -----TCTGTGTCAATGACCCGAGACCGGGCCGAACAGCGCGAG 715  
 OY 598 GCGTCCGTCACGAGGGGCTGCGCGTGTCCATGACGTCCTCGATGACCGGCGCGAGAAC 657  
 Db 716 CAGGCGGCGAAGGAGGTGTCACTTACTTCAAGCTGTGTGAGCGCGCGCCGACCGAG 775  
 OY 658 CCGCTCGAAAATGACGTTTGACGATGCTGTGCTTCAAGGCCGAGCGGACGCGACGAGCTG 717  
 Db 776 CCGGCGACGACCTGCTGTCCGCGCTTATCAGGGTCCAGAGACGACGATGACGCTGGGCTC 835  
 OY 718 AGCAGGAAGAGCTGTGTGCGCGCTGTGGGTGCGATTAATCGCTGCTGTGCAACCGATACAGC 777  
 Db 836 AGCGCCGACGAGCTGTCTCTCATGCGGTGTGTGTGCTGTGCGCCGGTTTGAAGGCGTGC 895  
 OY 778 ATCTACCTTATCGAGGTTGCTGTGTCTCAACTGTGTCGCGGTGCGCCGAGGCGCTCAGAGTG 837  
 Db 896 GTGAGCTTCATCGGGATGTGGACCTACCTGCTGTCTCAACCCAGCCGAGACAGCTCCGCTG 955  
 OY 838 GTGAAGGCGAGCGCGGGCTCATAGAGAACCGCTCGATGAGGTGTCTCCGCTTGCACAAAT 897  
 Db 956 GTGCGGCGGAGACCGGTGCGCGCTGCCCAACCCCGTGGAGAGATCTCGCTACATCCCT 1015  
 OY 898 ATCTCAAGAAATAGGAATCTGTGCTTTTGGCCAGGACGACCTGAGAGTAATGGGGGGCATCG 957  
 Db 1016 CTGCAGAGAC---CACCAACGCGCTTGCCTCCGAGGAGACTGTGAGATGTGGCGGTGTGCGG 1077

QY 958 ATCAAGAAAGGAGATGCTTCTTCTCGATCCGAGGCCCTTGAGATGAGACTGA 1017  
DB 1073 ATCCCCCAATACAGCAGGCTGCTGTCGAAAGCGCGCCCAACCGCAACCGCAAGCAG 1132  
QY 1018 TTCTCCAGGACCAAGCTGTTTATGTCACAGGAGCAGGAGCTTCGGTACGCT 1077  
DB 1133 TTCCGAGACCCCAACGCTCCAGCTCAACCGGAGAACCGCGCACTCTGTGTCGG 1192  
QY 1078 AAGAGCCCCATGCTGCCCCGGGGGTGTCCTTTCGCTCGAGCGAGATCGCGTG 1137  
DB 1193 CAGGGCATCTCCTGTCATGAGCGCGCGCTGGCCAGCTGAGGAGGAGGTGGCGTG 1252  
QY 1138 GGCACATCTCTCGATGTTCCCGAGATGAAGCTG 1173  
DB 1253 CGGGCCTGTTGGCGCGCTTCCCGCTGTGCGCTG 1288

RESULT 6  
ABX56047  
ID ABX56047 standard; DNA; 1347 BP.

AC ABX56047;

DT 13-FEB-2003 (first entry)

XX M. echinospora calicheamicin biosynthesis gene calW.

XX Calicheamicin biosynthetic gene cluster; aryltetrasaccharide;  
KM aglycone; calicheamicin resistance; nonchromoprotein enediylne;  
KM enediylne resistance; bone marrow cell; gene therapy; gene; ds.

OS Micromonospora echinospora spp. calichensis.

PN WO200279465-A2.

XX 10-OCT-2002.

XX 28-NOV-2001; 2001WO-US44285.

XX 28-NOV-2000; 2000US-0724797.

XX (SLOK ) SLOAN KETTERING INST CANCER RES.

XX Thorson J;

XX WPI; 2003-092897/08.

XX P-PSDB; ABU11985.

XX Novel nucleic acid molecule from nonchromoprotein enediylne biosynthetic  
PT gene cluster from Micromonospora echinospora useful for conferring  
PT calicheamicin resistance on a subject -

XX Claim 9; Page 120-122; 179pp; English.

XX The present invention relates to the isolation of the Micromonospora  
CC echinospora spp. calichensis calicheamicin biosynthetic gene cluster  
CC encoding proteins and enzymes used in the biosynthetic production of  
CC calicheamicin, including aryltetrasaccharide and aglycone. The gene  
CC cluster also includes the gene encoding for the protein that confers  
CC calicheamicin resistance. The calicheamicin biosynthetic gene cluster  
CC is a nonchromoprotein enediylne biosynthetic gene cluster. Expression  
CC vectors containing genes from the biosynthetic gene cluster are also  
CC disclosed. The expression vectors are useful for producing calicheamicin  
CC biosynthetic proteins. The calicheamicin self-resistance gene provides  
CC an approach for gene therapy, for example, by introduction of enediylne  
CC resistance genes into bone marrow cells, thus increasing resistance and  
CC allowing tolerance to chemotherapeutic doses of calicheamicin.  
CC ABX56028-ABX56073 represent genes from the M. echinospora calicheamicin  
CC biosynthesis gene cluster.

XX Sequence 1347 BP; 163 A; 505 C; 513 G; 166 T; 0 other;

XX Query Match 10.2%; Score 128.8; DB 25; Length 1347;

Best Local Similarity 49.7%; Pred. No. 4,2e-20;  
Matches 356; Conservative 0; Mismatches 357; Indels 3; Gaps 1;

QY 459 CCCGATGCGCGGATCAAGCGCTCTGTGTAAGTTCCGCGAGTGTACAGAACTTCCG 518  
DB 558 CCGTGGTCCGTGTTCAAGTATCCGTCGCCGCGGTGATCTCGAATCTCGCGCT 617  
QY 519 TCGCTTGGCTCGGAGACTGCGCGCGCTCGCGCGGTGTTGGTCCCGAGTGTGA 578  
DB 618 GCGGTACCGGAAACAGCGGTTCTTGAAGCGCGGTCAACCGGATCTGATGCGCGCT 677  
QY 579 GAGACCAAGACCTGTGTGCGCTCCGTCAACGAGGAGGCTCGCGCTCTCATAGAGCTCT 638  
DB 678 GCGCGCGAGACAGCGCGCGCGCGGTGCGCGCGAGATCCGCGAGTCTCTGAGCGGTGT 737  
QY 639 CGATAGCGCGCGAGAAACCGCTCGGAAATAGCTCTTGAAGATGCTCTTCAAGCCGA 698  
DB 738 GACCGACAAGAGCGCGCACCC---CGGCGACGACGTGCTCAAGCCGCTGCTCGCGCGCA 794  
QY 699 GCGCGACGCGACGAGCTGAGCAGAGAGAGCTGCGCGCTGCGGTGGGATTAATGCG 758  
DB 795 GCGCGCGCGCGCGAGCCCGACACAGAGCGGTGTGAGCATGCTCTGTCTGTCTGT 854  
QY 759 TCGTGCACCGATACACAGATCTTATGCGCTGCTGCTCAACCTGCTGCGCTC 818  
DB 855 GCGCGGACGCTCAAGAGCTGCAATATGATCTGCTGAGCGTGTGCGCTGTGACCA 914  
QY 819 GCGCGAGCGCTCGAGCTGTGTGAAGCGCGCGCGCTCATGAGAAACGCGCTGTGATGA 878  
DB 915 TCGGAGCGCGCTGCGCGCGCTGCGCGAGCGGACCGGTCTCCCGCGCGCTGAGGA 974  
QY 879 GGTGCTCGCGCTTCAATATCTCAGAAATGGAATGCGGTTCGCGAGCAGAGACT 938  
DB 975 GCTGCTGCGGTACTTACCATGTGTGAGCGCGAGCCGCGGACCGCGAGCT 1034  
QY 939 GAGTACTGCGCGGATCGATCAAGAAAGGAGATGCTCTTCTCTGATCCGAGCGC 998  
DB 1035 GACGCTCGGTGGGTCAACATCCGCGCGGAGGAGGTGTGTGGGCTGTGGCAGGCGC 1094  
QY 999 CCTGAGAGATGGAGATGTTATCTCAAGGTCAGACGTGTTTATGTGCGAGGAGCAAGG 1058  
DB 1095 CAACCGGAGCGCGCGCGGTGTGACCGGCGGACGATTCGACCGCGAGCGCGCGC 1154  
QY 1059 CGCGAGCTCGCGGTAGGTAGAGGCCCATGTGCGCGCGGTGCTGTGCTGCGCT 1118  
DB 1155 GCACCACTGCGCTTGTGCTGACGAGACATGTCGCCGCGAGCACTGCGCGCT 1214  
QY 1119 CGAGCGGAGATCGCGGTGGGACCACTTCCGTAGTGTCCCGAGATGAAGCTGA 1174  
DB 1215 GGAATCGACGTCGCGGTGAGCGCGGTGTGCGCGGTGCGCGGTGCGGTGA 1270

RESULT 7  
AAA75635  
ID AAA75635 standard; DNA; 5970 BP.

AC AAA75635;

DT 22-JAN-2001 (first entry)

XX Nucleotide sequence of ORF12 which encodes a transcriptional activator.

XX Nardanolide synthase; polyketide synthase gene; nardanolide polyketide;

XX antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;

XX desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;

XX picromycin biosynthesis; ss.

XX Streptomyces venezuelae.

XX US6117659-A.  
XX 12-SEP-2000.



XX 28-MAY-1998: 98US-0087080.  
 PR 28-AUG-1998: 98US-0141908.  
 PR 22-SEP-1998: 98US-0100880.  
 PR 08-FEB-1999: 99US-0119139.  
 XX  
 PA (KOSA-) KOSAN BIOSCIENCES INC.  
 PI Ashley G, Belach MC, Belach M, McDaniel R, Tang L;  
 DR WPI: 2000-072618/06.  
 DR P-PSDB: AAY67209, AAY67213, AAY67218.  
 XX  
 PT New recombinant DNA encoding a domain of narbonneide polyketide  
 PT synthase, for production of ketolide antibiotics -  
 XX  
 PS Disclosure; Page 35-37; 98pp; English.  
 XX  
 CC This is contig 002 from the recombinant cosmid PKOS023-27 DNA sequence  
 CC (see AA556001) which contains a Streptomyces venezuelae DNA insert. The  
 CC cosmid contains open reading frames which encode the various modules of  
 CC the narbonneide polyketide synthase (PKS). The invention relates to  
 CC recombinant DNA containing a coding sequence for a narbonneide PKS.  
 CC Polyketides are compounds synthesized from 2-carbon units through a  
 CC series of condensations and subsequent modifications. Modular PKSs are  
 CC responsible for the production of many antibiotics including picromycin.  
 CC The narbonneide PKS consists of a loading module, six extender modules,  
 CC and two thioester domains. Four proteins make up the narbonneide PKS  
 CC (PICAI, PICAIL, PICAIIV and PICAIIV). PICAI includes the loading module  
 CC and extender modules 1 and 2, PICAIIV includes extender modules 3 and 4,  
 CC PICAI includes extender module 5 and PICAIIV includes extender module 6  
 CC and a type II thioesterase domain. The second type II thioesterase  
 CC domain is found on the PICB protein. The nucleotide sequences encoding  
 CC all of these proteins can be isolated in recombinant form from the  
 CC recombinant cosmid PKOS023-27. Narbonneide is desosaminylated in S.  
 CC venezuelae to yield narbonneide, the desosaminyl transferase enzyme is  
 CC required for this conversion, and the desosaminyl biosynthetic genes are  
 CC also found in cosmid PKOS023-27. The recombinant DNA of the invention is  
 CC used to express, in transformed cells, narbonneide (or its derivatives)  
 CC or other ketolides (particularly hybrids), which may then be converted.  
 CC (e.g. by other enzymes recombinantly expressed in the same hosts) to  
 CC polyketide antibiotics or their intermediates. The antibiotics are useful  
 CC in human or veterinary medicine.  
 XX  
 SQ Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;  
 Query Match 9.8%; Score 123; DB 21; Length 5970;  
 Best Local Similarity 47.5%; Pred. No. 1.2e-18;  
 Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;

QY 596 TCGGCTCCGTCACCGAGGGGCTCGGCTGCTCCATGACGTCTCTGATGAGCGGCGCAGGA 655  
 DB 1924 -----CCATGGCCCGAATGAGCGGCTATCTCTCCCGCTCATCGACTCCAGCGGCGC 1977  
 QY 656 ACCCGCTGAAATACGCTCTTGAAGATGCTGCTTCAAGCGCGAGCGCAGCGCAGCGC 715  
 DB 1978 AGGACGGCGAGAGACCTGCTCAGCGGCTCTGTCGCGGACAGCGAGCGAGCGCTCCGCGC 2037  
 QY 716 TGAGCAGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775  
 DB 2038 TCACCTCCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2097  
 QY 776 CGATCTACCTTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 835  
 DB 2098 CGGTCAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2157  
 QY 836 TGGTGAAGGCGAGCGGCGCTCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 895  
 DB 2158 CCTCGGCGGCGAGCATGAGCGCTTGTGAGCGGCGGCTGAGAGAGATGTTGCGCTACGAGG 2217  
 QY 896 ATATCTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 955  
 DB 2218 GCCCGTGAATCGCGGAGCTTACCGCTTCCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2277  
 QY 956 CGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1015  
 DB 2278 TCATCCCGCGGCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2337  
 QY 1016 TATCTCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1075  
 DB 2338 GCTTCGCGGAGCGGAGCGGCTTGTGAGCGGCGGCTGAGCATCGCGGAGGAGGAGGAGGAGG 2397  
 QY 1076 GTAGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1135  
 DB 2398 GCCAGGAGATCACTTGTGAGCGGCGGCGGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGG 2457  
 QY 1136 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1186  
 DB 2458 TCGGCGGCTTGTGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2508  
 RESULT 9  
 ID ABS56092 standard; DNA; 5970 BP.  
 XX  
 AC ABS56092;  
 XX  
 DT 21-JAN-2003 (first entry)  
 XX  
 DE Contig 002 from cosmid PKOS023-275 containing S. venezuelae DNA insert.  
 XX  
 DE Narbonneide polyketide synthase; PKS; desosaminyl biosynthetic gene;  
 KW desosaminyl transferase gene; beta-glucosidase gene; antibiotic;  
 KW pick hydroxylase gene; C12 hydroxylase gene; narbonneide;  
 KW desosaminylated polyketide; narbonneide biosynthesis; mutant;  
 KW picromycin biosynthesis; de.  
 XX  
 OS Streptomyces venezuelae.  
 OS Synthetic.  
 OS  
 PN WO200297062-A2.  
 XX  
 PD 05-DEC-2002.  
 XX  
 PF 22-FEB-2002; 2002WO-US05642.  
 XX  
 PR 22-FEB-2001; 2001US-0793708.  
 XX  
 PA (KOSA-) KOSAN BIOSCIENCES INC.  
 PI Ashley G, Belach MC, Belach M, McDaniel R, Tang L;  
 XX



DR WPI: 2003-041412/03.  
 XX Preparation of polyketides by recombinant DNA technology, useful as  
 PT antibiotics and as intermediates in the synthesis of pharmaceutical  
 PT compounds -  
 XX  
 PS Disclosure; Page 40-41; 127p; English.  
 XX  
 CC The present invention relates to recombinant DNA sequences encoding  
 CC for a naphthalene polyketide synthase (PKS) domain, and methods of  
 CC producing polyketides by recombinant DNA technology. The recombinant  
 CC DNA sequences are derived from Streptomyces venezuelae desosamine  
 CC biosynthetic, desosaminyl transferase, beta-glucosidase, or pick  
 CC (C12) hydroxylase genes. The method is useful for transforming a cell  
 CC with a recombinant expression vector that encodes a functional  
 CC beta-glucosidase gene, and therefore for increasing the yield of a  
 CC desosaminylated polyketide in a cell. The recombinant methods and  
 CC materials are useful for expressing polyketides with significant  
 CC antibiotic activity, derived in whole or in part from the naphthalene  
 CC PKS gene, and other genes involved in naphomycin and picromycin  
 CC biosynthesis in recombinant host cells. The present sequence  
 CC represents contig 002 from cosmid pKOS023-2/5 containing  
 CC S. venezuelae DNA.  
 CC  
 SQ Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;  
 Query Match 9.8%; Score 123; DB 25; Length 5970;  
 Best Local Similarity 47.5%; Pred. No. 1.2e-18;  
 Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;  
 QY 236 AGTACTCTCGGCAATTCGAGCTCAGCATATGAGAAAGTATGTTGCGGCTGC 295  
 DB 1582 ACTCCAGACATCCCTGACCGAGCGCGGCTCAACCAACATGCTGAGTCCG 1641  
 QY 296 CGCGGAGATCAGCTCTGGGTCGCAAGCTGTCAACCCGCTTTACGTCACCGGCA 355  
 DB 1642 ACCGCGCGGACACCGCGCTGCGCAAGCTGTGCGCCGCTGATGATCAATGCGCGG 1701  
 QY 356 TCGACCTGCGGCGCGGCAATACAGGCGCATGCGTGTGATGCTGCTGCTCG 415  
 DB 1702 TCGAGTGTCTGGGCGCGGCTCAGAGATGTGAGGCGGTCTGTGACGCGATGCTGG 1761  
 QY 416 GACAAGAGAGTTCAGCTTGTGCGGATTAACGCGAGGAAATCCGATGCGCGGATCA 475  
 DB 1762 CGGCGCGCGAGCGCGCGCCATGTAGAGATCCCTGCGCGCTGC---CGATCA 1818  
 QY 476 GCGCTGTGTAAAGTTCGCGCGAGTGTGACGAGAGTTTCGCTTCCGCTCGCGGA 535  
 DB 1819 CCGTGTATCTCCGAATCTCTCGCGCGTCCGAGCCGAGCCGCGCTTCCGCGTGTGA 1878  
 QY 536 CTGGCGCGCGCTGCGCGGTGTGAGTCCCGAGGTGATGAGAACCAAGACCTTGG 595  
 DB 1879 CCGAGCGCTTGTGCTTCCCGAGCATCCCGCCAGGCCCAAGACG----- 1923  
 QY 596 TCGCGTCCGTACCGAGGCGCTCGCGCTGTCCATGACGTCTCCATGAGGCGCGCAGA 655  
 DB 1924 -----CGATGCGGAGATGAGCGGCTATCTCTCCGCTCATGATCCAAAGCGCGGCG 1977  
 QY 656 ACCGCTCGAAATGACGTCTTGAAGATGCTTCAAGCGCGAGGCCAGCGCAGACGC 715  
 DB 1978 AGGACGCGAGAGACCTGCTCAGCGCGCTGTGTCGAGCAGACGAGAGCGGCTCCGCG 2037  
 QY 716 TGAGACGAGAGAGTGTGCGCGCTCGCGGTGATGATGCTGCTGCGACCGATACCA 775  
 DB 2038 TGAACCTCGAGAGTGTGCTGATGAGGCCCATCTCTGCTCGCGGCGCAGAGACCA 2097  
 QY 776 CGATCTACCTTATCGCTGCTGCTCAACTGCTCGGCTCGCGGAGGCGCTCGAGC 835  
 DB 2098 CGGTCAATCTGATGCGCAACGCGCATGTAGCGCTGTCTCGAGACCCGACAGCTGCGCG 2157  
 QY 836 TGTGTAAAGCGGCGCGGCTCATGAGAGAACGCGCTGATGAGGTGCTCGCTTGCACA 895  
 DB 2158 CCTGCGGCGGACATGACGCTCTTGAACGCGCGGTGAGAGAGATGTTGGCGCTACGAG 2217

QY 896 ATATCTCAGAAATAGAACTGTGCTTTCGCGAGCAGACCTGAGTACTGCGGCGAT 955  
 DB 2218 GCCGCGGTGAAATCGCGACCTACGCTTCCCGGTGAGGCCGTGACCTGAGACGGACCG 2277  
 QY 956 CGATCAAGAAAGGAGATGCTTTTCTCTGATCCGAGCGCCCTGAGAGATGGACTG 1015  
 DB 2278 TCATCCGCGCGGAGACACGCTTCTGCTGCGCGAGCCGACCGCACCCCGGAGC 2337  
 QY 1016 TATCTCCAGGCGCAGCTGTGATGAGAGGAGCAGCGGCGGAGCCTCGGCTAGC 1075  
 DB 2338 GCTTCCGAGACCGACCGCTTCAACATCGCGCGGAGACCGCGCCATCTCGCTTCCG 2397  
 QY 1076 GTAGAGCCCCCATGTCTGCGCGGAGTGTCTTGTCTGCTGCGCTCGAGCGAGATCGCG 1135  
 DB 2398 GCCACGGATCACTTCTGATCGGCCCGCTTGGCCCGGTTGAGAGCCCGGATCGCG 2457  
 QY 1136 TGAGCCACCATCTTCCGATGAGTTCGCGAGATGAAGTGAAGAAACTCCCG 1186  
 DB 2458 TCCGCGCCCTTCTGAAACGCTGCCGAGCTGCGCTGAGCTCTCCCG 2508

RESULT 10  
 ABN88910  
 ID ABN88910 standard; DNA; 1191 BP.  
 XX  
 AC ABN88910;  
 XX  
 DT 21-AUG-2002 (first entry)  
 XX  
 XX Streptomyces sp. cytochrome P450 encoding DNA SEQ ID NO.1.  
 DE  
 XX Streptomyces sp. cytochrome P450, anticancer; UCN-01; monooxygenase;  
 KW Streptomyces; cytochrome P450; anticancer; UCN-01; monooxygenase;  
 KW microorganism; gene; ds.  
 XX  
 OS Streptomyces sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1191  
 FT /tag= a  
 FT /product= "cytochrome P450"

XX JP2002058490-A.  
 XX  
 XX 26-FEB-2002.  
 PD  
 XX 22-AUG-2000; 2000JP-0251304.  
 PP  
 XX 22-AUG-2000; 2000JP-0251304.  
 PR  
 XX (KYOW ) KYOWA HAKKO KOSYO KK.  
 PA  
 XX WPI: 2002-410015/44.  
 DR P-PSDB; ABB81315.  
 DR  
 PT Novel cytochrome P450 used for improving the productivity of anticancer  
 PT drug UCN-01 -  
 XX  
 XX Claim 5; Page 9-11; 22pp; Japanese.  
 PS  
 XX The present invention describes 3 Streptomyces cytochrome P450 (1)  
 CC sequences, ABB81315 to ABB81317 encoded by ABN88910 to ABN88912  
 CC respectively. (1) have monooxygenase activity and improve the  
 CC productivity of an anticancer drug designated UCN-01. (1) is used  
 CC for the production of UCN-01 with transformed Streptomyces sp.  
 CC microorganisms.  
 CC  
 SQ Sequence 1191 BP; 169 A; 423 C; 431 G; 168 T; 0 other;

Query Match 9.7%; Score 122.6; DB 24; Length 1191;  
 Best Local Similarity 48.3%; Pred. No. 1.1e-18;  
 Matches 430; Conservative 0; Mismatches 434; Indels 27; Gaps 2;

Oy	298	CCGAGGAGATCAACGCTCCGGGTCGCGAAGCTCGTCAACCCCTCTTTATCGTCAACGCGCATC	357
Db	271	CCGCCCCAGACACCCCGCTGGCGGGGCTGATTCGCCCGGGAATTCACCGGACGCGCGGCTC	330
Oy	358	GACCTGTCCGCGCGCGAATAACAGCGCACCGTCGCACAGCTGCTCGATCTCGCTCCGGA	417
Db	331	GAGCGCGCTCCGCGCCAGAGGCTCCAGAGAGATCACCGATCTCCTCTGGAACAGATGTGCGCG	390
Oy	418	CAAGAGGAGTTGAGCGTTGTGGGGGATTACGGCGGAGGGAAATCCCATGTGGCGGATCACG	477
Db	331	CGCGGCGCGCGCGACCTGCTGATCGAGTCTTTCGGGTACCCGCTGCCCTCACCGTCATCTCG	450
Oy	478	GCTCTGTGAAGTTTCCGGCCGAGGTGTGACGAGAAATTCCTGCGCTTTCGACTTCGCGACT	537
Db	451	GAACTCTCTCGGGGTGCCGAGA-----TTCACCGGGGGGCGC	486
Oy	558	GCGCGCGCGCTCCGGCGGTGGTTGTGTCGCCCAAGTTCGATAGAGAGACCAAGACCTCGTGC	597
Db	487	TTCCGCAAGCTGTCCAGCGAGGCGGGTGGACCAACAGCGGGGAGAGCCAGATAC--GGG	543
Oy	598	GCGTCCGTACCGAGGGGCTCCGCTGCTCCATGACGTCCTCGATAGCGCGCGCAGGAAC	657
Db	544	GCTTTCGTTCACATCGCGCGCTTACCTCGAGAGAGTTGTGAGAGAGAAAGCGGTGCGCACCG	603
Oy	658	CCGCTCGAAAATGACGTCTTGAACGATCTGCTTCAGGGCCGAGGCGGACGCGACAGGCTG	717
Db	604	CCGGCGGACGATCTGCTGACGCGCTGATCCGAGACGACGAGAGGACGCGGACCGCTCTG	663
Oy	718	AGCAGGAAGAGACTGATCGCGCTCGTGGGTGGATTAATGCTGCTGGACCGATACACAG	777
Db	664	TCACCCGCGGAGCTCGCGGACATGCGCTTATCTGCTATGCGCGGACACAGACACAC	723
Oy	778	ATCTACCTTATGCGGTTGCTGTGCTCAACTGCTGCGGTGCGCCCGAGGCGCTCGAGCTG	837
Db	724	GTCACACTTCATCACCGGTCCGCTCCACGCGCTCTTCACCCGATCCCGGCAACTCGGCCAG	783
Oy	838	GTCGAGGCGCGAACCCGGGCTCATATGAGGAACGGGCTCGATGAGAGTGTCCGCTTCGACAT	897
Db	784	GTCGCGGGGAGATGAGCTTCGTGGAACGCGGCTGTGTGAAGAGACGTGTGCGCATGAGGGG	843
Oy	898	ATCTCTCAGAAATGAGAACTGTGCGTTTCGCGCAGGACAGGACTTGAGATGACTGCGGGGATCG	957
Db	844	CCGGTGGAGAAACGCGACGTTCCGCTTCGCGCGCGAGCGGCTGTGAGATGAGGGGCAACGATC	903
Oy	958	ATCAAGAAAAGGGAGATGCTTTTCTCTCTGATTCGCAAGCGCCCTGTGAGAGATGGAACTGTA	101
Db	904	ATCCACAGCGGGGAGACCCGGTGTCTGATCGGGGCTGTGACCGCGCGGATGAGGACGCGCGCCG	963
Oy	1018	TTCTCTCAGGACAGACGTTGTTATGTGCACAGGAGACACGGGCGCGAGCTTCGCGTACGCT	107
Db	964	TATCCCGGCGCCGACCGCTTGACATCCACCGGGACACACGGGGGATCTCGGCTTTCGCG	102
Oy	1078	AGAGGCGCCCATGTCTGCGCCCGGGGTGCTCCCTTGTCTGCGCTTCGAGGCGGAGATGCGCGTG	113
Db	1024	CACGGCATTCACCTTCTGCTTCGCGCGCGCGCTTCGCGCGCTGTGAGGCGCGGGTGTGCTTG	108
Oy	1138	GGCACACATTTCCGTAAGTTTCCCGCAGATGAAGCTGAAGAAATCCGCCGTG	1188
Db	1084	CGCGGCTCTGTGAGCGGTGCTGTGTCTTCACGCGCGGACCGGGGACCGGGG	1134

RESULT 11  
AAZ87301  
ID AAZ87301 standard; DNA; 1251 BP.  
XX  
XX AAZ87301;  
AC  
XX  
DT 05-JUN-2000 (first entry)  
XX  
DE  
XX  
S. venezuelae macrolide biosynthetic gene p1kC, SEQ ID NO:38.  
XX  
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin  
KM neomethylmycin; narbomycin; polyhydroxyalkanoate monomer synthase;

KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
KV chronic obstructive pulmonary disease; respiratory inflammation;  
KM hypercholesterolaemia; crop protection agent; ds.  
OS Streptomyces venezuelae ATCC15439.

XX Location/Qualifiers  
FH Key 1..1251  
FT CDS /tag= a  
FT /product= "Pik"  
FT /transl\_except= (pos:307..309, aa:Leu)

PN MO20000620-A2.  
XX  
XX  
PD 06-JUN-2000.  
XX  
PF 25-JUN-1999; 99WO-US14398.  
XX  
PR 26-JUN-1998; 98US-0105537.  
XX  
PA (MINU ) UNIV MINNESOTA.  
PI Sherman DH, Liu H, Xue Y, Zhao L;  
PI WPI: 2000-160679/14.  
DR P-PSTD; AAY77196.  
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.,  
PT synthesis of methymycin and pikromycin -  
XX  
PS Claim 15; Page 428-429; 438pp; English.

The invention relates to an isolated and purified nucleic acid segment comprising a desosamine biosynthetic gene cluster, a fragment or its biologically active variant, where the nucleic acid sequence is not derived from the eryC gene cluster of Saccharopolyspora erythroa or Streptomycetes antibiotics. The invention also relates to a macrolide biosynthetic gene cluster, or fragments thereof. The macrolide biosynthetic gene cluster encodes proteins which synthesise methymycin, pikromycin, neomethymycin, narbomycin or a combination of these compounds. Recombinant or augmented cells comprising the desosamine and/or macrolide biosynthetic gene clusters are useful for the production of biologically active macrolides. The macrolide biosynthetic proteins are useful for synthesis of methymycin, pikromycin, neomethymycin and narbomycin. The alternative termination of these synthetase may be useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA) monomers. The compounds produced by the recombinant host cells are useful as biopolymers, e.g., in packaging or biomedical applications, to engineer PHA monomer syntheses or to prepare biologically active agents, such as chemotherapeutics, immunosuppressants, agents to treat asthma, chronic obstructive pulmonary disease as well as other diseases involving respiratory inflammation, cholesterol-lowering agents or macrolide-based antibiotics which are active against a variety of organisms, e.g., bacteria, including multi-drug resistant pneumococci and other respiratory pathogens, as well as viral parasitic pathogens, or as crop protection agents (e.g., fungicides or insecticides) via expression of polypeptides in plants. Sequences AAZ87295-287302 represent macrolide biosynthetic genes from Streptomyces venezuelae ATCC 15439, which encode proteins

AA77190-Y77197.

XX  
XX Sequence 1251 BP; 171 A; 484 C; 413 G; 183 T; 0 other;

SQ Query Match 9.6%; Score 121.4; DB 21; Length 1251;  
Best Local Similarity 50.4%; Pred. No. 2,1e-18;  
Matches 295; Conservative 0; Mismatches 299; Indels 0; Gaps 0

OY 602 CCCTGACCGAGGGGTCTGCCTTCATGACGTCCTTGATGAGCGGCCGACAACCCG 661  
DB 569 CCATTGGCGCATGAAGAAGCGGGTACTCTCCCGGCTCATGCACTCCAAGCGCGGGAGAGC 628  
OY 662 TCQAATAATGACGTTTGAGAGTGTGCTTCAAGCGCGGACGAGGACGAGGCTGAGCA 721  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 639 GCGAGGACCTGCTCAGCGCGCTGTCGCGACGAGCAGAGACGCGTCCCGGCTGACCT 688
Qy 722 CGAAGAGCTGCTCGCGCTCGTGGGATGATATCGTGTGCGACCGATACCATGATCT 781
Db 689 CCGAGGACCTGCTCGCTGATGCGCCACATCTCTGTCGGGGGACAGACACGATCTCA 748
Qy 782 ACTTATGCGCTGCTGCTGCTCAACCTGCTGCGGTCGCGCCGAGGCGCTGAGCTGTA 841
Db 749 ATCTGATCGCCAAACGCGATGTAACGCGCTGCTGCTGACACCCGACAGCTGGCCGCTGC 808
Qy 842 AGCGGAGCGCGGCTCATAGAGAAACGCGCTGATGAGTGTCTCGCTTCCACATATCC 901
Db 809 GGGCCGACATGACGCTTGTGAACGCGCGGGTGAAGAGATGTTGGCTTACGAGGCGCCG 868
Qy 902 TCAGATAGAACTGTGCTTTCGCGAGGACGAGCCTGAGATGCGGGGATGATCA 961
Db 869 TGGAAATCGCGACCTTACCGCTTCCGGTCCGAGCCCTGATGAGACGCGATCTCC 928
Qy 962 AGAAGGGAGATGCTTTCTCTGATCCGAGCGCCCTGAGAGATGGACTGATTTCT 1021
Db 929 CGCGCGGTGACAGCGGTCTCTGCTGCTGCGCGACCGCACCGCCGAGCGCTTCC 988
Qy 1022 CAGGCGCAGAGCTGTTTATGATGCGACGCGGACGCGGCGGACCTCGCTGATAGAG 1081
Db 989 CGAACCACCGACCTTTCACATCCCGCGGACACCGCGCCATCTCGCTTCCGCGCACG 1048
Qy 1082 GCGCCCATGCTGCGCGCGGGGTGCTTGTGCTGCGCTGAGGCGGAGATTCGCGGAGCA 1141
Db 1049 GCATTCACCTTCTGCACTCGCGCGCCCTTGGCCCGGTTGAGGCGCGGATCGCGCG 1108
Qy 1142 CCATCTTCCTGATGCTTCCCGAGATGAACTGAAAGAAACTCCCG 1186
Db 1109 CCTTCTCGAAGCGTGGCCGAGACCTCGCCCTGGAAGTCTCCCGG 1153

```

## RESULT 12

ABZ76672  
ID ABZ76672 standard; DNA; 1992 BP.

AC ABZ76672;

DT 30-APR-2003 (first entry)

DE Streptomyces sp. TM-7 related compactin DNA SEQ ID NO.1.

XX Streptomyces sp. TM-7; pravaestatin; compactin; hyperlipidaemia;

KM antilipaeamic; microorganism; gene; ds.

OS Streptomyces sp. TM-7.

XX Key Location/Qualifiers

FT CDS 544..1761

FT /\*tag= a

FT /product= "compactin related protein (ABP95987)"

FT /tag= b

FT /product= "compactin related protein (ABP95988)"

XX MO20029109-A1.

XX 12-DEC-2002.

XX 30-MAY-2002; 2002MO-JP05252.

XX 01-JUN-2001; 2001JP-0166412.

XX (SAOC) MERCIAN CORP.

XX Fujii T, Hirose S, Aritoku Y, Morimiya T, Johdo O, Isehiki K;  
WPI; 2003-148672/14.  
XX P-PSDB; ABP95987, ABP95988.

PT Novel Streptomyces sp. produced polypeptide for hydroxylation of  
compactin at 6beta-position and its encoded DNA, applicable in  
PT constructing transformant microbes to synthesize pravaestatin for  
treating hyperlipidaemia

PS Claim 1; Page 45-48; 67pp; Japanese.

XX The present sequence represents a DNA sequence which contains a base  
CC sequence from bases 544-1758 in the sequence of (1) with 1992 base pairs,  
or a DNA hybridisable with the DNA under stringent conditions and  
CC encoding a polypeptide with hydroxylase activity on compactin at  
CC 6beta-position. Also described: (1) DNA containing base sequences from  
CC bases 544-1758 and from bases 1782-1970 in the sequence of (1) or a DNA  
CC hybridisable with the DNA under stringent conditions and encoding a  
CC polypeptide with hydroxylase activity on compactin at the 6beta-position;  
CC (2) a polypeptide encoded by any of the DNA or containing an amino acid  
CC sequence based on the polypeptide but with some amino acids deleted,  
CC substituted or added and having hydroxylase activity on compactin at  
CC the 6beta-position; (3) a recombinant DNA obtained by integrating with  
CC any of the DNA; (4) a microorganism transferred with the recombinant DNA;  
CC (5) a process for producing pravaestatin by culturing the transformant  
CC microorganism before isolating the culture liquor or cells, and addition  
CC of compactin for reaction to give pravaestatin for recovery; and (6)  
CC Streptomyces sp. TM-6 (FERM BP-8002) or TM-7 (FERM BP-8003). (1) has  
CC antilipaeamic activity. The polypeptide and its encoded DNA are applicable  
CC in constructing transformant microorganisms to synthesize pravaestatin for  
CC treating hyperlipidaemia. With the recombinant microorganisms,  
CC pravaestatin can be produced efficiently, with much less galpha  
CC hydroxylated epimer formed.

XX Sequence 1992 BP; 275 A; 691 C; 726 G; 300 T; 0 other;

Query Match 9.6%; Score 120.8; DB 25; Length 1992;

Best Local Similarity 50.7%; Pred. No. 3.1e-18; Matches 290; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

```

Qy 602 CCGTACCCAGAGGCGCTCGCGCTGCTCCATGACGTCCTGATGAGCGGCGAGAACCCGC 661
Db 1112 CCGCCCGGAGAGACTGAGAGCTATCTGGCGCTGCTGTGAGACGCTGGAGCGAGT 1171
Qy 662 TCGAAATGACGCTTTCGATGCTGCTTTCAGGCGGAGCGCGACGAGGCTGAGCA 721
Db 1172 CCGCGCGGCGCTCTGTCGACGCTGTCAGAGGAGAGCTGAGAGGCGCGATCGAC 1231
Qy 722 CGAAGAGCTGTCGCGCGCTGTCGATGATTCGTCGTCGTCGTCGTCGTCGTCGTCG 781
Db 1232 GCGAGGACCTGTGTGCGCGCGATCTCTGCTGTGTGTGCGGACACAGCGCGAGCT 1291
Qy 782 ACTTATGCGCTGCTGCTGCTCAACCTGTCGCGTCCCGAGGCGCTCGAGCTGTGA 841
Db 1292 CGATGACGTCGCTGAGGTATCAACCTCTCGAACAATCCGACACAGACCCCGGTTC 1351
Qy 842 AGCGGAGCGCGGCTCATGAGAAACGCGCTGATGAGTGTCTCGCTTCCACATATCC 901
Db 1352 GCGCGGATCCGCTGCTGTCGCGCGCGGTGAGGAGACTCTGCGCTATCTGCGCATCG 1411
Qy 902 TCAGATAGAACTGTGCTTTCGCCAGGACGACCTGGAATACGCGGCGCATGATCA 961
Db 1412 CCGACATGCGCGGCGGAGATCGGACGCGGACATGAGATGACGCGGACGCGCATCC 1471
Qy 962 AGAAGGGAGATGCTTTCTCTGATCCGAGACGCGCTGAGAGATGAGACTGATTTCT 1021
Db 1472 GGGGGGGGAGGGGTATCTCATCAACATGATGATGCAACCGGACGCGCTCGCTTCCG 1531
Qy 1022 CAGGCGCAGACGCTGTTGATGTCGACGCGGACGCGGCGGACGCTCGCTGATAGAG 1081
Db 1532 CCGACCCGAGAGCTTTCGACGTCGCGGCGGAGGCGGACACCACTGCGCTTCCGCTACG 1591
Qy 1082 GCGCCCATGTCGCGCGGGGTGCTCTGCTGCTGCTGAGGAGGAGATGCGCGGAGCA 1141
Db 1592 GGGTGCATCACTGCTCGCGGAGAACTGCGCGCGCTGGAAGGTCTATCTTCACGG 1651
Qy 1142 CCATCTTCCTGATGCTTCCCGAGATGAAGCTG 1173

```

```

Db      1652  CCCTGTTGAGCGGCTGCCGCTCTCGGCTG 1663
RESULT 13
ABV75558/c
ID      ABV75558 standard; DNA; 36538 BP.
XX
AC      ABV75558;
XX
DT      22-JAN-2003 (first entry)
XX
DE      Saccharopolyspora butenyi-spinosyn biosynthetic gene cluster 2.
XX
KW      Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;
XX      metabolite; spinosyn; gene; ds.
XX
OS      Saccharopolyspora sp.
XX
PH      Key
CDS     Location/Qualifiers
FT      complement (114..938)
FT      /tag= a
FT      /product= "bush"
FT      /note= "No start codon given"
CDS     1389..2561
FT      /tag= b
FT      /product= "bush"
FT      /tag= c
FT      /product= "bush"
FT      complement (3359..4546)
FT      /tag= d
FT      /product= "bush"
FT      4684..6303
FT      /tag= e
FT      /product= "bush"
FT      6317..7510
FT      /tag= f
FT      /product= "bush"
FT      7555..8406
FT      /tag= g
FT      /product= "bush"
FT      /note= "No start codon given"
FT      8640..9572
FT      /tag= h
FT      /product= "bush"
FT      /note= "No start codon given"
FT      complement (9668..10666)
FT      /tag= i
FT      /product= "bush"
FT      complement (10675..12135)
FT      /tag= j
FT      /product= "bush"
FT      complement (12864..14177)
FT      /tag= k
FT      /product= "bush"
FT      14627..15970
FT      /tag= l
FT      /product= "bush"
FT      /note= "No start codon given"
FT      16008..17144
FT      /tag= m
FT      /product= "bush"
FT      /note= "No start codon given"
FT      17168..17917
FT      /tag= n
FT      /product= "bush"
FT      complement (18520..19932)
FT      /tag= o
FT      /product= "ORF LI"
FT      complement (19978..20488)
FT      /tag= p
FT      /product= "ORF LI"

```

```

FT      /note= "No start codon given"
FT      complement (20536..21033)
FT      /tag= q
FT      /product= "ORF LIII"
FT      /note= "No start codon given"
FT      21179..21925
FT      /tag= r
FT      /product= "ORF LIV"
FT      complement (22671..23453)
FT      /tag= s
FT      /product= "ORF LVI"
FT      complement (23687..24886)
FT      /tag= t
FT      /product= "ORF LVII"
FT      complement (26177..26923)
FT      /tag= u
FT      /product= "ORF LVIII"
FT      /note= "No start codon given"
FT      27646..28476
FT      /tag= v
FT      /product= "ORF LIX"
FT      /note= "No start codon given"
XX
PD      WO200279477-A2.
XX
PD      10-OCT-2002.
XX
PF      28-MAR-2002; 2002WO-US09968.
XX
PR      30-MAR-2001; 2001US-280175P.
XX
PA      (DOWC ) DOW AGROSCIENCES LLC.
XX
PI      Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
XX      Mitchell JC;
XX
DR      MPI: 2003-058434/05.
XX
DR      P-REDB; ABB57683, ABB57684, ABB57685, ABB57686, ABB57687, ABB57688,
XX      ABB57689, ABB57690, ABB57691, ABB57692, ABB57693, ABB57694,
XX      ABB57695, ABB57696, ABB57697, ABB57698, ABB57699, ABB57700,
XX      ABB57701, ABB57702, ABB57703, ABB57704.
XX
PT      New butenyl-spinosyn biosynthetic genes, useful for increasing the
XX      production of butenyl-spinosyn insecticidal macrocides, or for changing
XX      the metabolites or products produced by spinosyn-producing
XX      microorganisms
XX
PS      Claim 2; Page 99-119; 218pp; English.
XX
CC      The invention relates to a novel DNA molecule comprising a DNA sequence
XX      that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
XX      PKS (polyketide synthetase) domain, or a spinosyn PKS module. The
XX      butenyl-spinosyn biosynthetic genes are useful for increasing the
XX      production of butenyl-spinosyn insecticidal macrocides. The genes are
XX      also useful for changing the metabolites or products produced by
XX      spinosyn-producing microorganisms. The present sequence represents a DNA
XX      molecule encoding butenyl-spinosyn biosynthetic enzymes.
XX
SQ      Sequence 36538 BP; 6867 A; 12266 C; 11182 G; 6223 T; 0 other;
XX
Query Match      9.5%; Score 119.6; DB 25; Length 36538;
Best Local Similarity 48.2%; Pred. No. 9.6e-18;
Matches 424; Conservative 0; Mismatches 429; Indels 27; Gaps 2;
QY      291 GCTGCCGCGGAGATCAAGCTCGGTCGCAAGCTGTCACCGCTGTTACGTCACG 350
DB      2462G GATGACCCCGCGGAGATCAAGCTCGGTCGCAAGCTGTCGTAAGCGGTTACCGCACG 24567
QY      351 CGCCATCGACTCTCGCGCGCGCAATACAGCGACCGTCGACCGAGCTCTGATGCTCG 410
DB      24566 TCGAGTCGAAACAGCTTCGCGCCCGCACCCAGAGCTGTTGACCACTTACGTCGAT 24507
QY      411 CTCGGGACAAAGAGATTGACGTTGTGGGGGATTACGCGAGGGAATCCGATCGCGCG 470

```

Db	24506	GGCCGCGGAGCGGGCCGCTTGAGCGACCTGATGTCAGGCGCTGGCCCTGCGCTGCATCA	24447
Qy	471	GATCAGCGCTCTGTGTAAGTTCCGGCCGAGTGTACGAGAGTTCCGTGCTTGCGCTC	530
Db	24446	GGTGTATTCGACACTTGTGGGGCAATTCGAGTGGCCGACCCGACGCTTCCGGGTCTGCTC	24387
Qy	531	GGCGACTGCGCCGCGCGCTCGCGCGTGGTGTGGTGCCCAAGTGTGATGAGAGACCAAGAC	590
Db	24386	GGACATTGTCGTGGCGATCACTGAAACAGTCTGTGAGAGATTCCGGGAAAGCCGCGATCA	24327
Qy	591	CTGTGTGCGCGTCCGTCACGAGGGGGCTCGCGTGTCTCATGACGTCTGTGATGAGCGCGG	650
Db	24326	AATTCCGCGGTATTCGAGAACTGGTCC-----AGAGCG 24291	
Qy	651	CAGGAAACCCTCGMAAATGACGTTCAAGATGCTGCTTCAAGCCGAGGCGACGCGAG	710
Db	24290	CAGAAATATCCCATCTAGAGACCTGTGACGCGTGTGTGTCTGAGGCCAGAGCGGAAGAC	24231
Qy	711	CAGGCTGAGCAGAGAGAGCTGTGTCGCGCTGTGGGTGCGATTATGCTGTGACCGA	770
Db	24230	CCAGCTAGCGAGGAGAGATTGTGTCACCCGCGCGGGTCTGTATCGCGGGTTGCA	24171
Qy	771	TACCAAGATTAACCTTATCGCGTTCCGTGTGCTCAACTGTCTGCGGTCCGCCGAGCGCT	830
Db	24170	GACCACCGCTAACACATCCGAACCTTCAACTGTCTCAACCTGCTCAACCATCCGAGACAGCT	24111
Qy	831	CGAGCTGTGAAGGCGAGCCGGGGCTCAAGAGAAAGCGCTCGATGAGATGCTCCGCTT	890
Db	24110	TGACAACTGATGCGCGATCAAGCTGTGGTCCGCGCGGTGAGAGATTTCTGGCTA	24051
Qy	891	CGACAAATATCTCT--CAGATATAGAACTGTGCGCTTTCGCGAGGACGACTGGAATCTGA	947
Db	24050	CAGCGCGCTGTGGTCCACACTCTGGTTCCCGCGCATCTGCACTGAGGATCTGMACTGGG	23991
Qy	948	CGGGGCAATCGATCAAGAAAGGGAGATGTCCTTCTCTTGATCCGAGCGCCCTGAGAGA	1007
Db	23990	CGAGTGTGCAATTCGCGCGGGGTGACGAGATGTTCTTTCGAGATGCTCGCGCAATCTGGA	23931
Qy	1008	TGGGACTGTATTTCTCCAGGCGCAACGTGTTGATGTGTCGACGGGACACGGGCGGACCTT	1067
Db	23930	CAGTGCCTGTTCGACCGGGCGGAGCGMACTGACCTGTGCTGGGAGACACAACTCGCACT	23871
Qy	1068	CGGCTAGCGTGAAGGCCCGCATGTGCCCCGGGGGTGTCCCTTGCTGTGCTCTGAGGCGGA	1127
Db	23870	GGGCTTGGGSCAGGCGCCGCACTACTGCACTGGCGGCCCAATTGGCGAGATGAGACTGCA	23811
Qy	1128	GATGCGCGTGGGCAACATCTTTCGTGAGATTCCCGAGATG 1167	
Db	23810	GGTAGCGATCGGCACTGATCAACGGTTTTCCGCACTGG 23771	

RESULT 14	
ABN86911	
ID	ABN86911 standard; DNA; 1449 BP.
XX	
AC	ABN86911;
XX	
DT	21-AUG-2002 (first entry)
XX	
DE	Streptomyces sp. cytochrome P450 encoding DNA SEQ ID NO:3
XX	
KM	Streptomyces, cytochrome P450; anticancer; UCN-01; monooxy
XX	
microorganism; gene; ds.	
XX	
OS	Streptomyces sp.
XX	
FH	Location/Qualifiers
FT	1..1449
FT	/*tag= a
XX	
FT	/product= "cytochrome P450"
XX	
FN	JP2002058490-A.

XX 26-FEB-2002.  
PD  
XX  
XX 22-AUG-2000; 2000JP-0251304.  
PF  
XX  
XX 22-AUG-2000; 2000JP-0251304.  
PR  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
XX  
XX WPI: 2002-410015/44.  
DR  
DR P-PSDB; ABB81316.  
XX  
XX Novel cytochrome P450 used for improving the productivity of anticancer  
PT drug UCN-01 -  
XX  
XX  
XX Claim 5; Page 12-14; 22pp; Japanese.  
PS  
XX  
XX The present invention describes 3 Streptomyces cytochrome P450 (I)  
CC sequences, ABB81315 to ABB81317 encoded by ABB88910 to ABB88912  
CC respectively. (I) have monooxygenase activity and improve the  
CC productivity of an anticancer drug designated UCN-01. (I) is used  
CC for the production of UCN-01 with transformed Streptomyces sp.  
CC microorganisms.  
XX  
XX  
SQ Sequence 1449 BP; 231 A; 620 C; 390 G; 208 T; 0 other;

Query Match 9.4%; Score 118.6; DB 24; Length 1449;  
Best Local Similarity 48.5%; Pred. No. 9.3e-18;  
Matches 417; Conservative 0; Mismatches 409; Indels 33; Gaps 2

QY	305	ATCAGCGTCGGGATCCGGAGAAGTCGTCAACCCGTCGTTACGTCACCGCCATATCAGACTGCG	364
Db	521	AACACGCCCGGCTACGGGGCGCTGGTGAACGAAGGGTTTCAACAACGGGGGCGGTGGCGGCTG	580
QY	365	TGCGCGCCGAATAACAGGACACCGTCGACCACTGTCGTGATGCTCTGCCAGCAAGAAG	424
Db	581	TGCCCCCTTACATCTCTCTCCCTGGTGCAGAGTTGCTCGAACACTGGCCGAACTCCAGNA	640
QY	425	AGTTGGAAGTTGTCGCGGATTTAGCGCGAGGGGAATCCGGAATGCGGCGCGATTCAGCGCTCTGT	484
Db	641	CGGTGAACCTGATGCGCGAAGCTCCGGGGCCGCTCCCGTACGGTCAATCTTGCAGAGCTGC	700
QY	485	TGAAGTTCCGGCCGAGATGTGACAGAAATTCGCGTCTGTCGAGCTCGGCGACCTGGCCGCG	544
Db	701	TCGGGGGTGCGGGAATTC-----GAAACGGCGGCTG	728
QY	545	CGCTCGCGGTGCGTTTGGTGTGCCCAAGTCGATGAGAGAACCAAGACCTTGGTCGCTCG	604
Db	729	CGTACGCAACCTGTGTGACGC-ACCTGTTCTGTCGCGGAGAACCCGCAAGCGATTCGACGCTCG	787
QY	605	TCACCGAAGGGGCTGCGCGCTGCTCATATACGTCCTCGATGAGAGGGGCGAGAAACCCGCTCG	664
Db	788	CTTCCCAAGCCGTGGGCGAATTACATGACCGGCTCTGTGCGCGCAAGGCGAACCCGACCCG	847
QY	665	AAATGACGTCCTTGACGATGCTGCTTGAAGCCGAGCCGACGCGACGAGCTGAGACAGA	724
Db	848	GCGACACGCTGCTGACGACCTCATTCGCGGTACGACGACGAGAACCAACTGTGTCGAGG	907
QY	725	AGGAGCTGTCGCGCTGTGTGGTGCATTAATGCTGTGCGAACCGATATCCAGATCTAAC	784
Db	908	ACGAACCTGTATCCCTCGCCGTACTCTGTGCTGTGGCGGCGACAGACACACCAACT	967
QY	785	TTATTCGGTTTCGCTGTCGTCAACTGTCGCGGTCGCGCCGAGGCGCTCGAGCTGGTGAAG	844
Db	968	TCATTCGCAACGCGCCCTTGCGCTTCTTACGCGAACCGGAATCGCTTCGCGCACTTGAAGG	1022
QY	845	CCGAGCCCGGAGCTCATGAGGAACGCGTCTGATGAGTGTCTCGCTTCGACAATATCTCA	904
Db	1028	CCGAGCCGCAACTCTCGGCGGCGCACTGACGAGTTGCTCGCTGATCAAGACTGCCCCGTG	1082
QY	905	GAAATGGAATGTGTCGTTTGCCACGAGCAGAACCTGGAATATCTGGGGGCAATCATCAAG	964
Db	1088	GGATTCGCAAGTTCGCTTACGACACGAGAGGCTTCAAGCTTCGGCGGCAACGAGATCTCCG	1147

QY 965 AAGGGAGATGCTTTCTCTGATCCGAGCGCCCTGAGAGATGAGATCTATTCTCCA 1024  
DB 1148 AAGGGATACCGGCTCCATCGACCGCGCGCCCAATCGACCGGATTCCTCCG 1207  
QY 1025 GGCACAGCGCTGTTGATGTCGACGAGGACCGGCGCGAGCTCGGATGAGAGGCC 1084  
DB 1208 ACCCGGACCGCTGAGCTCACCACCGGCGGCGCACCGGTCACTGGCTTCGGGACAGCA 1267  
QY 1085 CCCATGCTGCCCCGGGGGTGCTCTGTCGCTCGAGCGGAGATCGCGTGGGACCA 1144  
DB 1268 TCCACCGCTGCGGAGGCGCCGTTGGCCGAGCCGACCGGAACTGGCCCTTCACGCGC 1327  
QY 1145 TCTTCGTAGTTCCTCCGA 1163  
DB 1328 TCATCACCGCTACCCCA 1346

RESULT 15  
ABX56043  
ID ABX56043 standard; DNA; 1209 BP.  
XX  
XX ABX56043;  
AC  
XX  
XX 13-FEB-2003 (first entry)  
DT  
XX  
XX M. echinospora calicheamicin biosynthesis gene calS.  
DE  
XX  
XX Calicheamicin biosynthetic gene cluster; aryltetrasaccharide;  
KW aglycone; calicheamicin resistance; nonchromoprotein enediyne;  
KW enediyne resistance; bone marrow cell; gene therapy; gene; ds.  
XX  
XX  
OS Micromonospora echinospora spp. calichensis.  
PN  
XX WO200279465-A2.  
XX  
XX 10-OCT-2002.  
PD  
XX  
XX 28-NOV-2001; 2001WO-US44285.  
XX  
XX 28-NOV-2000; 2000US-0724797.  
PR  
XX  
XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
PA  
XX  
XX  
PI Thorson J;  
DR WPI; 2003-092897/08.  
DR P-PSDB; ABU11981.  
XX  
XX  
PT Novel nucleic acid molecule from nonchromoprotein enediyne biosynthetic  
PT gene cluster from Micromonospora echinospora useful for conferring  
PT calicheamicin resistance on a subject -  
XX  
XX  
PS Claim 9; Page 110-111; 179pp; English.  
XX  
XX The present invention relates to the isolation of the Micromonospora  
CC echinospora spp. calichensis calicheamicin biosynthetic gene cluster  
CC encoding proteins and enzymes used in the biosynthetic production of  
CC calicheamicin, including aryltetrasaccharide and aglycone. The gene  
CC cluster also includes the gene encoding for the protein that confers  
CC calicheamicin resistance. The calicheamicin biosynthetic gene cluster  
CC is a nonchromoprotein enediyne biosynthetic gene cluster. Expression  
CC vectors containing genes from the biosynthetic gene cluster are also  
CC disclosed. The expression vectors are useful for producing calicheamicin  
CC biosynthetic proteins. The calicheamicin self-resistance gene provides  
CC an approach for gene therapy, for example, by introduction of enediyne  
CC resistance genes into bone marrow cells, thus increasing resistance and  
CC allowing tolerance to chemotherapeutic doses of calicheamicin.  
CC ABX5602-ABX5673 represent genes from the M. echinospora calicheamicin  
CC biosynthesis gene cluster.  
XX  
XX Sequence 1209 BP; 132 A; 451 C; 452 G; 174 T; 0 other;

Query Match 9.3%; Score 116.8; DB 25; Length 1209;  
Best Local Similarity 47.7%; Pred. No. 2.3e-17;  
Matches 488; Conservative 0; Mismatches 512; Indels 24; Gaps 4;  
QY 128 ACTGGATGAAAGGCGCTCTTGGAGTCTCACCCTGATACAGACGTGTGGGTTC 187  
DB 101 ACCGTGGGCCACACGGGGCTTGTATGTCCTCCGCTTGCGGACGTGCGGGGTCTCG 160  
QY 188 GGGAGAAACGCTTCGCGGTCACTGAGAGAGTGGAAATCGAGCCGGAAGTACTGTGG 247  
DB 161 GGGAGAACGCTTCGCGGACCGGACCGGATCGCGGTTGAGACCTGCTGGGCGC 220  
QY 248 CCATTCGCGAGCTCAGCATATGAAAGATACGATGTCGAGTGTGGGTGCGCGGAGATC 307  
DB 221 GGCT-----GCTGCCAGATCTGCGGACATATCTCTGTCAGAGACGACCCGACC 274  
QY 308 ACGCTCGGGTCGCAAGCTGCTCAACCCGCTGTTTACGTCAAGCGCCATCGACTGTGTC 367  
DB 275 ACGGCGGGGTGGGGGGGTGGGTGGGCGCGGCTTCTGCGCGGTGGCGCGGCTGG 334  
QY 368 GCGCGGAAATACAGCGCACCGTTCACACAGCTGCTGATGCTGCTCGGACAAAGAGT 427  
DB 335 AACCGGTATCGCCCGACACCGTGAACGACTGTCGCGCCCGCTGCGCCGGGCGCA 394  
QY 428 TCGAGCTTGTGGGATTACGCGGAGGAAATCCGATGCGCGGATCAGCGCTGTGTA 487  
DB 395 TGGAGCTGTGTCAGACGAGCTGGGCTACCGGCTGGCGCTGCGCGGTCTCTGGCTGTG 454  
QY 488 AGGTTCCGCGCGAGTGTGACGAGAAATTCGTGCTTGGCTGGGCACTGCGCGCGC 547  
DB 455 GCTCCGCGCGCGCACTGGGGGGGTGGGCGCT-----GTTGCGCGGACG 502  
QY 548 TCGGCTGGGTTTGTGTCGCCAGCTGATGATGAGAGAACCAAGCCCTGTCGCTCTCA 607  
DB 503 TGGAGCGAACCTTGACCGGGGCGCGCAAGCGCGGAGACATGCGCGGACGCGGCGA 562  
QY 608 CCGAGGGGCTGCGGCTGCTCCATGACGTCTGATGAGCGGCGAGAACCGGCTCGAAA 667  
DB 563 TGGCGGATTCGCGCACTACGTGAGCGGGCTCTGCGAGGCGGGCGGTGAGGGCGCG 622  
QY 668 ATGAGCTTTGACGATGCTGCTTCAAGCCGAGCGGACGCGAGCGTGAAGCAAGG 727  
DB 623 AGGACTGTGGCGTTGATGCTGACGCCCA---GACCGCGGCTGATGATGTCGCAAG 679  
QY 728 AGCTGTGCGGCTGTGGGTGATTAATGCTGCTGCTGCGACCGATACAGATCTACCTTA 787  
DB 680 AGATGTGACGACGGGTGATCAGTTCACTTTCACCGGCGACGAGACGGTGGCAGCGAG 739  
QY 788 TCGGCTGCTGTGCTCAACCTGCTGCGGTCGCGCGCGCGAGCGCTCGAGCTGTGAAGCGC 847  
DB 740 TGGGCAAGCGCGTGTGAGCTGTGAGCGGACCGGACCGGACGACTGACTCGGGCGCC 799  
QY 848 AGCGCGGCTCATGAGAAACGCGCTGATGAGTGTCTCGCTTGCACAAATATCTCAAGA 907  
DB 800 GCGCGGACCTGTGCGCCAGAGGCGCTGAGAGTGTCTCGGTACGACCCGTCGTCGAGT 859  
QY 908 TAGGAATGTGCTTTCGCCAGGACGACCTGGAATATCGCGGGCATGATCAAGAAAG 967  
DB 860 CCAACACCCCGGAGCTGAC---GTGCGATGAGACTGCGGCTGGCGGCTGCGCGCG 916  
QY 968 GGGAGATGCTTTTCTCTGATCCGAGCGCCCTGAGAGATGGAATGATATTCACAGC 1027  
DB 917 ACGAGCTGTGTGCTCTGCTGCGCGCGCGGCGGACACCGGACCCGCGGGGTACGACCGC 976  
QY 1028 CAGAGCTTTGATGTGCGACGAGACGAGGCGGAGCTTCGCTGACGTGAGAGCCCC 1087  
DB 977 CCGAGATTTGACATCAGACGAGGATTCGGTCCCTCATGTCTTTCGCGCGGGGATGTC 1036  
QY 1088 ATGTTCGCGCGGGGTGCTCTTGTGCTGCTGAGAGCGGAGATGCGCGGTGACATCT 1147  
DB 1037 GCTACTGCTTGGGTCTTACTCTGCGCCGATCGACACTGGCGCGCGGTGGCTGG 1096  
QY 1148 TCCG 1151

Db 1097 CCGG 1100

Search completed: October 3, 2003, 12:28:58  
Job time : 317.318 secs

---



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:41 ; Search time 4507.41 Seconds

(without alignments)  
11435.853 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_62369\_63628

Perfect score: 1260

Sequence: 1 atgacacagagagcagcgaa.....agccctccaagctcgatag 1260

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sgs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sgs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_gy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	100.0	16124	6 AX024277	AX024277 Sequence
2	1260	100.0	68750	6 AF210843	AF210843 Sorangium
3	1260	100.0	68750	6 AR193029	AR193029 Sequence
4	1260	100.0	68750	6 AR199551	AR199551 Sequence
5	1260	100.0	68750	6 AR199559	AR199559 Sequence
6	1260	100.0	68750	6 AR199567	AR199567 Sequence
7	1260	100.0	68750	6 AR201097	AR201097 Sequence
8	1260	100.0	68750	6 AR208671	AR208671 Sequence
9	1246.2	98.9	58733	1 AF217189	AF217189 Sorangium
10	1246.2	98.9	71989	6 AR172664	AR172664 Sequence
11	157.4	12.5	302650	1 AP005958	AP005958 Bradyrhiz
12	154	12.2	3078	1 SERCP450A	M83110 Saccharopol
13	149.4	11.9	1233	6 AX697983	AX697983 Sequence
14	149.4	11.9	60196	6 AX697977	AX697977 Sequence
15	149.2	11.8	66808	1 SNU421825	AJ421825 Stigmatel
16	148.8	11.8	301925	1 AP005046	AP005046 Streptomy
17	146.4	11.6	78210	1 AB070949	AB070949 Streptomy
18	146.4	11.6	300425	1 AP005022	AP005022 Streptomy
19	143.8	11.4	1257	6 AX097457	AX097457 Sequence
20	143.8	11.4	14759	1 AE004755	AE004755 Pseudomon
21	141.8	11.3	10167	1 SFU08223	U08223 Streptomyc
22	132	10.5	24268	1 AB070947	AB070947 Streptomy
23	132	10.5	300550	1 AP005030	AP005030 Streptomy
24	131.6	10.4	12420	1 AF521896S2	AF521895 Streptomy
25	130	10.3	2243	1 SERERYFGH	M54983 Saccharopol
26	128.8	10.2	1347	6 AX573690	AX573690 Sequence
27	128.8	10.2	90348	1 AF497482	AF497482 Micromono
28	127	10.1	298550	1 AP005029	AP005029 Streptomy
29	123.6	9.8	17512	1 AB071405	AB071405 Lechevall
30	123.6	9.8	25681	1 SAE414559	AJ414559 Saccharot
31	123.6	9.8	26144	1 AB090952	AB090952 Lechevall
32	123.6	9.8	28654	1 AF534707	AF534707 Lechevall
33	123.4	9.8	325483	1 AP005050	AP005050 Streptomy
34	123	9.8	1470	1 AF087022	AF087022 Streptomy
35	123	9.8	4342	1 AF079139	AF079139 Streptomy
36	123	9.8	5970	6 AR271682	AR271682 Sequence
37	123	9.8	5970	6 AR277666	AR277666 Sequence
38	123	9.8	32634	1 AF293355	AF293355 Streptomy
39	122.6	9.7	1191	6 BD133544	BD133544 Novel cyt
40	121.8	9.7	299925	1 AP005039	AP005039 Streptomy
41	121.2	9.6	69644	1 AY179507	AY179507 Streptomy
42	120.6	9.6	210614	1 AB088224	AB088224 Streptomy
43	120.4	9.6	6629	1 AB072568	AB072568 Streptomy
44	120.4	9.6	19601	1 AB070944	AB070944 Streptomy
45	120.4	9.6	299425	1 AP005049	AP005049 Streptomy

#### ALIGNMENTS

RESULT 1  
LOCUS AX024277 16124 bp DNA  
DEFINITION Sequence 82 from Patent DE19846493.  
ACCESSION AX024277 AX024384  
VERSION AX024277.1 GI:10184551  
KEYWORDS  
SOURCE  
ORGANISM  
Polyangium cellulosum  
Polyangium cellulosum  
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
Sorangineae; Polyangiaceae; Polyangium.

REFERENCE  
AUTHORS Beyer, S. and Mueller, R.J.  
JOURNAL Patent: DE 19846493-A 82 13-APR-2000;  
BIOTECHNOLOG FORSCHUNG GMBH (DE)

COMMENT On Oct 15, 2002 this sequence version replaced gi:10184588.  
 FEATURES  
 source  
 1.16124  
 /organism="Polyangium cellulosum"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:56"  
 BASE COUNT 2336 a 5400 c 5695 g 2693 t  
 ORIGIN  
 Query Match 100.0%; Score 1260; DB 6; Length 16124;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-199;  
 Matchés 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 ATGACACAGAGGACGACGATGAGTGAAGGACGCTTTCGACTTCAAGCCGTTTC 60  
 Db 7086 ATGACACAGAGGACGACGATGAGTGAAGGACGCTTTCGACTTCAAGCCGTTTC 7145  
 Qy 61 GGGCCCTGGGTAGCGCGAGAGACCGGTTCCCGCGATGAGCGCGCTGAGAGGACCCGCC 120  
 Db 7146 GGGCCCTGGGTAGCGCGAGAGACCGGTTCCCGCGATGAGCGCGCTGAGAGGACCCGCC 7205  
 Qy 121 ATCTTCTAGGATGAGGACGCGCTCTGAGTCTTCAAGGATCCAGACGCTGTCAGCG 180  
 Db 7206 ATCTTCTAGGATGAGGACGCGCTCTGAGTCTTCAAGGATCCAGACGCTGTCAGCG 7265  
 Qy 181 GTGTTCCGCGAGACGCTTCCGCGTCACTCCAGAAAGATGGGAATCGACCGCGAGTAC 240  
 Db 7266 GTGTTCCGCGAGACGCTTCCGCGTCACTCCAGAAAGATGGGAATCGACCGCGAGTAC 7325  
 Qy 241 TCGTGGGACATTCGCGAGCTCAGCGATGAGAAAGATGTTTCGAGGCTGCGCGCG 300  
 Db 7326 TCGTGGGACATTCGCGAGCTCAGCGATGAGAAAGATGTTTCGAGGCTGCGCGCG 7385  
 Qy 301 GAGGATCAGCGCTCGGGTCCGCAAGCTGTCACACCCGCTTTACGTCACGCGGCATCGAC 360  
 Db 7386 GAGGATCAGCGCTCGGGTCCGCAAGCTGTCACACCCGCTTTACGTCACGCGGCATCGAC 7445  
 Qy 361 CTGCTCGCGCGCGAATTAAGCGCCACCGTTCAGTCACTGCTGCTGCTCCGAGCA 420  
 Db 7446 CTGCTCGCGCGCGAATTAAGCGCCACCGTTCAGTCACTGCTGCTGCTCCGAGCA 7505  
 Qy 421 GAGGATTCAGCGCTTGGCGGATTAAGCGGAGGAAATCCCGATGCGCGCATCAGCGCT 480  
 Db 7506 GAGGATTCAGCGCTTGGCGGATTAAGCGGAGGAAATCCCGATGCGCGCATCAGCGCT 7565  
 Qy 481 CTGTTGAAGGTTCCGCGCGAGTGTGACGAAAGTTTCGCTTCGCTGCGGACTGCG 540  
 Db 7566 CTGTTGAAGGTTCCGCGCGAGTGTGACGAAAGTTTCGCTTCGCTGCGGACTGCG 7625  
 Qy 541 CGCGCGCTCGCGTGGGTTTGGTGGCCCGAGTGTGAGAGAACCAAGCCCTGTCGCG 600  
 Db 7626 CGCGCGCTCGCGTGGGTTTGGTGGCCCGAGTGTGAGAGAACCAAGCCCTGTCGCG 7685  
 Qy 601 TCGCTACCGAGGAGGCTCGCGCTGCTCAATGACGCTCCTCGATGAGCGCGCAGAAACCG 660  
 Db 7686 TCGCTACCGAGGAGGCTCGCGCTGCTCAATGACGCTCCTCGATGAGCGCGCAGAAACCG 7745  
 Qy 661 CTTCGAAATGACGCTTTCGATGATGCTTCAGGCGCGAGCGCAGCAGCTGAGC 720  
 Db 7746 CTTCGAAATGACGCTTTCGATGATGCTTCAGGCGCGAGCGCAGCAGCTGAGC 7805  
 Qy 721 AGGAGGAGACTGTCGCGCTCGGCTGGGATTAATGCTGCTGCGACCGATACCAAGATC 780  
 Db 7806 AGGAGGAGACTGTCGCGCTCGGCTGGGATTAATGCTGCTGCGACCGATACCAAGATC 7865  
 Qy 781 TACCTATGCGCTGCTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
 Db 7866 TACCTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7925  
 Qy 841 AAGGCGAGCGCGGCTCATGAGAAACGCGCTGATGAGGCTGCTGCTGCTGCTGCTGCTG 900  
 Db 7926 AAGGCGAGCGCGGCTCATGAGAAACGCGCTGATGAGGCTGCTGCTGCTGCTGCTGCTG 7985

Qy 901 CTCAGATAGGAACTGTCGCTTTCGCGAGGAGACCTGAGTACTGCGCGGATCGATC 960  
 Db 7986 CTCAGATAGGAACTGTCGCTTTCGCGAGGAGACCTGAGTACTGCGCGGATCGATC 8045  
 Qy 961 AAGAAAGGAGATGCTCTTCTCTCTGATCCGAGCGGCTGAGAGATGGAATGATATTC 1020  
 Db 8046 AAGAAAGGAGATGCTCTTCTCTCTGATCCGAGCGGCTGAGAGATGGAATGATATTC 8105  
 Qy 1021 TCCAGCCAGACGCTTTCGATGTCGAGCGGACACGCGCGGAGCTTCCGCTGATG 1080  
 Db 8106 TCCAGCCAGACGCTTTCGATGTCGAGCGGACACGCGCGGAGCTTCCGCTGATG 8165  
 Qy 1081 GGGCCCAATGTCGCGCGGAGTCCCTGTCGCTCGAGCGGAGATGCGCGTGGGC 1140  
 Db 8166 GGGCCCAATGTCGCGCGGAGTCCCTGTCGCTCGAGCGGAGATGCGCGTGGGC 8225  
 Qy 1141 ACCATCTTCGATGCTCCCGAGATGAGAGCTGAAAGAACTCCCGTTCGATACAC 1200  
 Db 8226 ACCATCTTCGATGCTCCCGAGATGAGAGCTGAAAGAACTCCCGTTCGATACAC 8285  
 Qy 1201 CCCGCTTCGAGACATGATCACTCAAGCTCATCTTGAAGCCCTCCAAAGCTGATG 1260  
 Db 8286 CCCGCTTCGAGACATGATCACTCAAGCTCATCTTGAAGCCCTCCAAAGCTGATG 8345  
 RESULT 2  
 AF210843 68750 bp DNA linear BCT 21-JAN-2000  
 LOCUS  
 DEFINITION  
 Sorangium cellulosum strain so ce90 epothione biosynthesis gene cluster, complete sequence.  
 ACCESSION  
 AF210843  
 VERSION  
 AF210843.1 GI:6724237  
 KEYWORDS  
 ORGANISM  
 Polyangium cellulosum  
 Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.  
 REFERENCE  
 1 (bases 1 to 68750)  
 Molnar, I., Schnupp, T., Ono, M., Zirkle, R., Milmanow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D. D., Goriach, J., Mayo, J. M., Hu, A., Goff, S., Schmid, J. and Ligon, J. M. The biosynthetic gene cluster for the microtubule-stabilizing agents epothiones A and B from *Sorangium cellulosum* so ce90  
 Chem. Biol. 7 (2), 97-109 (2000)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 2 (bases 1 to 68750)  
 REFERENCE  
 Molnar, I.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (03-DEC-1999) Natural Product Genetics, Novartis Agribusiness Research Institute, Inc., 3054 Cornwallis Rd., P.O. Box 12557, Research Triangle Park, NC 27709, USA  
 FEATURES  
 source  
 1.68750  
 /organism="Polyangium cellulosum"  
 /mol\_type="genomic DNA"  
 /strain="so ce90"  
 /db\_xref="taxon:56"  
 <1.1826  
 /note="ORF1"  
 /codon\_start=3  
 /transl\_table=11  
 /product="unknown"  
 /protein\_id="AF26904.1"  
 /db\_xref="GI:6724238"  
 /translation="ASLDALPARATSARVLDDHGRATERHYLAERGIEDLARLREH LRIQGGSPFCHMCGLDLELLADQPLASIFHHASLRPDPSTPAMLVDPGLV RMLARGAPRIREYERERARITQERLMLAAAPCPAPDLPFRPDADAGPLGP MSPVLAERBRIRASATRPBELACALLAWICTGAGPMSGYAYEWLPENLILGFLPT ALAASAPETISDALRGARLPASWEVTSKKSQULNPEALMEBLRTIPAMGNADN LSRFERALIAEVRILRAQAPPAAGSLAVAGVSSSGRLISGLVTDALYSQDND IVMFQGRISPVLLAGTDPFEELAPLSOMLFVHANAGTISKVLTEGSPILIVARN QARPMSLVHARGFMAMVNOAMVPDERGAFVVGQSTIMEFHPTRCLHBPAGSAPS

CDS  
LACDEHLWCELSAGRLLEWHPHRRGAPRAYIGENPIAATWYPSLTNATHVL  
WADPRRAILADKRTGVEPIVLAETIRPPAHVVSIEDRIDIALTALGQSPSRDMVHEIR  
SADTVADYOROLMDRPDVLNRRGLFETTNDRILTLARS"  
complement (1900..3171)  
/note="ORF2; similar to hypothetical proteins from  
Mycobacterium tuberculosis and Streptomyces coelicolor and  
to D0-peptidases"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26905.1"  
/db\_xref="GI:6724239"  
/translation="MGALISVAACGALGAEESQPGQDAGALAPAREVMAEVA  
AGOMGAVMLVARGDVDAVGTLEGSAPMRSDTIFRIASMTKATATAMVLE  
EGKLDSPVDPRMLPELANRKYARIDGIDETVPAERPIVRDLMFTMGFGISPD  
SSPIORADEGLVNAQOPVMTPHGPDMEIRRLGTLPLMHOPGAMNTYSLVQGL  
VGRADQCFDAFVERRIIAPGMRDTHFVADKARPAAGCGFTDEGTETMDHD  
GAASVAVPAPPSGAGLVSTVDYLLFAELMNGVHEGRLLSASVREMTADHL  
TPAQKASSFPPEFETHGMGYGMAVVAIPDAVSEVPGRYGMDGFGTSMINDGRELI  
IGIWMQSAGFLPSGALERFWRSYVATESA"  
3415..5556  
/note="ORF3; similar to Na/H and drug/H antiporters"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26906.1"  
/db\_xref="GI:6724240"  
/translation="MHGCTEROVLSTVTLALVTNARASGELARLRQPVIGELRG  
GVVIGPSVGLAPGFRALRQEPNPAVGLSGISVIGLMLNLAIEVDGILRKEA  
RPGALSALGAIAPLAPAGAFSALVLDPLSGFLGLVLTAVSVAKVILIERESM  
RHSYAQVTLAAGVSEVAVMLVMTSSYGASPALAARSALASGLFLMTVYGR  
LTHAMRVADATRVSKGVSGLVLTFLAALQRLGHLPLGAFGLVTLNAPRT  
NRPLDGVQTLVAGLFAPVFLVAGMRYDSQLTPAMWGVALILTATATAKVPA  
LGALGCGRSEBALVANGLMKGTDLIVAVIGVELSLNEATYAVAVATVTRA  
SPALLMDEKAPPTQESARLERBEAKRAYITEVERILVIVAHALPGATUVES  
IVASKRKGEDVDTITELSEVQAPGSPRAAGSGLARLGVIGWRORLERGS  
IOAILRKSRLDLITIGARSPARAGMSFGLQDAIVQRAESNVLVVGDPAERAS  
ARLITVPIIGESYFAADLAHVLAHVAIDAEVLTLSSQTDPGAIVMREDRSPVRAV  
ARSVDAEVPFRGRUGVRSVSVHVGARPSDEITRELARAYDVLVIGCYHGLRL  
YLGSVSVVRSVPAVLLVYAHGTRBQVR"  
complement (5612..5992)  
/note="ORF4"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26907.1"  
/db\_xref="GI:6724241"  
/translation="MDKPIGTRCAIAEGYIPGSGNPEPQMTSHETACLINASDRDA  
QVATVFSDDPDAGPYVTVPARTRHRENDLTEPPIPRDTDYASVIESDAPIV  
OHTRLDSROAENALSTIATYDRE"  
6226..6675  
/note="ORF5"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26908.1"  
/db\_xref="GI:6724242"  
/translation="MKAYDTGRFRRRRTGHTLGLLALAGCGSPSEKTYQGTRLAP  
GADARVADVDPAATYTRLADVVLSPERLEAGSEFVVMQRPSPSPRRRGVLD  
YNADSRKGLAEITVYPANFELITAEKQSSPOSPPSSAAVIGPTSVG"  
7610..11875  
/gene="epoa"  
/note="epoa"  
/note="Epos A"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAF26919.1"  
/db\_xref="GI:6724253"  
/translation="WADRIERAAEDPIAIVGASCRLEPGVVIDLSEGMTLLSGSRDVI  
GRVPERWDAAMFDPDPADGKTPVTASRLSDVACPDASFPGISPEALRMDPAHR  
LLLEWCWALEMAAIAPASLVGTETGVFIGIGPSEYEMALPQATASAIIDAHGIGTGM

PSVAGRISVALGRPCVAVDTAYSSSLVAVHLACOSLSRGECSTALAGVSLMLSP  
STVUWAKTALADGCKAFSAEADDFGEGGCAVVVLRLESGARDGIRLIVGIG  
SAIWHDAASGLITVNGSSOEIVLKRLADAGCAASVYKAEAGTGLDPIEIOA  
LNAVYGRVAPFLLIGSVKTNLGHVEYSGITGLKVLISLQHGQIPATHIOALN  
PRISWGLRLITVTRAPTPWDMTPRAGYSFMSGKTNNAHVLEERPAATCTPABE  
RPAELVLARSALDOAQLARLDHLETPSOQLGVAASLATRAMEIRLAVAA  
SREGLRALDAAGQOTSPASVRSIAOSSKGLAFLTGOAQOTLGMGRGLYVMSAF  
REAFDLCVRLFNOELDRLEVMMAEASVDALDQTAQOPALPFEERGLAYLMS  
MGVEBELVAGHSITGELIACVAGVPGVADQOPVHAIANAARARTRKALHVSAPHS  
DVAAAVAPHAASVSIANAPOVIVAGADQPVHAIANAARARTRKALHVSAPHS  
PLAEMLEAFGRVAESVYRPSVILVNSLKGACTOEVSPGIVHAEVAVPAG  
VKALHAAGAGTPVEVGRKSTLGLVPCMDPARALLASRAQDEBATVLEALGGM  
AVGGLVWAGLIPSGGRVPLPTYPMQRERYWIDTKADDAAGDRAPAGHDEBEG  
GAVGGRSARLDHPPESSGREKVEAAGDRPRLIDEQVLDHILVYTERARPG  
LGEVEIVDAAGLSFNDVOLALGMVPLDQKPPAPRLIGGECGRIVAAGEVNGVY  
VGQVIALSAGAPRTHYTTSAALVLPAPOLSAIEAAMPAVATYATYALDRIARLOP  
GERVLIIAATGGVGLAIVQAQHVGAHVATAGIPEKAYLIESGIVYASDRDRFLR  
ADVAVATGGGVGLAVNLSLSELIDKSFNLLRSHGRVELEKRCVADNQGLRPLR  
NLSPSLVDLGMMLREPARVALIEELGILAGVFPPTATLPIARVADAFPSMQ  
AOHLGKLVTLGMDPEQIRIPTHAGAGPTGDRDLRLASAPAAALAEALRTQ  
VSOVLRTPEIKVGEALFTRLGMSLMAVELRNIEASLKLKSLTFLSTSPNIALLA  
ONILDLATATLSLERVAENLRAGVQNDPFSGADQWEIITAL"  
7640..8920  
/gene="epoa"  
/note="Region: beta-ketoacyl synthase/malonyl-CoA  
decarboxylase"  
9236..10201  
/gene="epoa"  
/note="Region: acetate acyltransferase"  
10529..11428  
/gene="epoa"  
/note="Region: enoyl reductase"  
11549..11764  
/gene="epoa"  
/note="Region: acyl carrier protein"  
11872..16104  
/gene="epop"  
11872..16104  
/gene="epop"  
/note="Epos P; contains condensation/heterocyclization,  
cysteine adenylation, putative NAD(P)H oxidase, and  
peptidyl carrier domains"  
/codon\_start=1  
/transl\_table=11  
/product="nonribosomal peptide synthetase"  
/protein\_id="AAF26925.1"  
/db\_xref="GI:6724259"  
/translation="MTINOLNLELHOGIKLADGERLQIOAPRNALPNILARISEH  
KSTILMLRQLPSESIVPAERHAFPLDIOESYWLQSGTGAFTVPSGIAHAREYD  
CTDLDPRLSRAPFKVVARHDMRAHLLPMMQVTEBKVDADIEIILRGDRSTRA  
RLVSLDRAMSHRYDTERPPLHYVAVNRDLERQRYLSDILINVDGSLSTIFKDM  
SYEDPETSLEBELSTRDYVLALBESRKSEBAHRSMDYMRRIAELEPPETPMKAD  
PSTLKEIRFRHTEQMLPDSWGRKRVGRGLTPVTIILAESEVIGRMSASRFTL  
NITLFNRLPVHPRVNDITGDTSMVLLDITTRKSEORAKRIOELWEMSDICVS  
GIEVREAAVVLGIORGALPEVVLITSLNQOVSVTLRGSTEVYSTOTPOLLDH  
OYVEDGDVLVANDIYDGVPPDLDDMLBLYVVLRLTEPGEBOVRSGLPAOLE  
ABASAKNTALLSHETLHGLFAARVQLPMQILAVSARKTLTYEELSRSSRLGARLY  
EQGARPTLVAVWMEKSEBOVAVLALLEGAAVYIPDADIPARITHYLDHGEYKLV  
LTOPDLGKLSWPEGIQRLVSEAGVGDQOPMPPIQTPSDLAIVITYSGSLPK  
GVMIDHGAANTLIDINERPEIGDGVNLASISPLSYDVDAIVYLAAGTIVPDA  
SKLADPAHMAELIEREKVTWNSVPALMRMLVEFBERPSLASLISLSDGMIYV  
GLPELOAIRGVVISLGGTASISGIVPVNVLSMSIPIYSGRLNRQTHVLD  
BALEPRVWVPGOLYTGCVLALGMSDEKTRSPVHPETGERLKYKTGLDGRVLD  
GNIEFMGRBNQITLGRYVBLGEIETLASHPVYRANVYIPVGNDAANULLAYVVP  
EGTRRAAEODASLKTERTIDARAAADAGLSDERQFKLARGLGRDLGKRVVDL  
TGQDPREAGLDVYARRSRVTFLEAPLPVEFGFSLCSLSEVDGATLPKFRPSAG  
STYEVQTYAVYKSGRIEGVDEGFYYYPFHRLILKSDHIGISCAHAROMQVDFDEAA  
FNLFVEGRIDAISLIGSSSRFCLLAGVMAOLIMQASGNCINGVONFONFEOYR  
PVLDLRSDVYVHMGICGRVDRPQFOVCTIGSCGRTATTRGAPRERBOHFAMLD  
FLTKLPEVYVPTVYVZELDLPITSNKKVRKALREKQDTPSHSGHTIROYDLBEI  
LVAAVREVLLEVAGLQOSFVDIGATSIHVIRMSLSLQKLDREIATITELFOYPNLDS  
LASGLRDSDSLQDPRMORVREVRKGRRS"  
16251..21749  
gene

```

CDS
    /gene="epob"
    16251..21749
    /gene="epob"
    /note="EPOS B"
    /codon_start=1
    /transl_table=11
    /product="polyketide synthase"
    /protein_id="AF26920.1"

Query Match      100.0%; Score 1260; DB 1; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.8e-199;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACACAGAGACGACGAAATCAGATGAGACGAAACCTGCTTCACTTCAAGCCGTTG 60
DB 62369 ATGACACAGAGACGACGAAATCAGATGAGACGAAACCTGCTTCACTTCAAGCCGTTG 62428
QY 61 GCGCCTGGGTAACGCGAGAGACCGCTTCCCGCGATGAGCGCTTGAAGAGCAACCCC 120
DB 62429 GCGCCTGGGTAACGCGAGAGACCGCTTCCCGCGATGAGCGCTTGAAGAGCAACCCC 62488
QY 121 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGCTCTGACCCGATACGACGAGTGGCG 180
DB 62489 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGCTCTGACCCGATACGACGAGTGGCG 62548
QY 181 GTGTTCCGCGACGACGCTTCCGCTGCTGATGAGAGAGTGGGAATCGACCGCGAGTAC 240
DB 62549 GTGTTCCGCGACGACGCTTCCGCTGCTGATGAGAGAGTGGGAATCGACCGCGAGTAC 62608
QY 241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTATGTTGGGGCTGCGCGCG 300
DB 62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTATGTTGGGGCTGCGCGCG 62668
QY 301 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGCTGTTACGTACCGCGCATGAC 360
DB 62669 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGCTGTTACGTACCGCGCATGAC 62728
QY 361 CTGCTCGCGCGCGCAATATACAGCGACCGCTGACACCTGCTGAGTCTGCTCCGACAA 420
DB 62729 CTGCTCGCGCGCGCAATATACAGCGACCGCTGACACCTGCTGAGTCTGCTCCGACAA 62788
QY 421 GAGGATTCGACGCTTGTGCGGAGTATACGCGAGAGGAAATCCGATGCGCGCATCAGCG 480
DB 62789 GAGGATTCGACGCTTGTGCGGAGTATACGCGAGAGGAAATCCGATGCGCGCATCAGCG 62848
QY 481 CTGTTGAAGGTTCCGCGCAAGTGTGACGAGAAATTCGCTGCTTCCGCTGCGGACTGCG 540
DB 62849 CTGTTGAAGGTTCCGCGCAAGTGTGACGAGAAATTCGCTGCTTCCGCTGCGGACTGCG 62908
QY 541 CCGCGCTCGCGCGTGGGTTTGTGCCCCAGGTGAGAGAGACCAAAACCTGCTGCGCG 600
DB 62909 CCGCGCTCGCGCGTGGGTTTGTGCCCCAGGTGAGAGAGACCAAAACCTGCTGCGCG 62968
QY 601 TCGTCACCGAGGCGCTCGCGCTGCTCAATGACGTCCTCGATGAGCGCGCAGAGAACCG 660
DB 62969 TCGTCACCGAGGCGCTCGCGCTGCTCAATGACGTCCTCGATGAGCGCGCAGAGAACCG 63028
QY 661 CTGCAAAATGACGCTTTGACGATGCTGCTTCAAGCCGAGCGCAGCGACGAGCTGAGC 720
DB 63029 CTGCAAAATGACGCTTTGACGATGCTGCTTCAAGCCGAGCGCAGCGACGAGCTGAGC 63088
QY 721 AGCAAGAGCTGTGCGCGCTCGGAGTATGAGTATGCTGCTGCGACCAATACCAAGATC 780
DB 63089 AGCAAGAGCTGTGCGCGCTCGGAGTATGAGTATGCTGCTGCGACCAATACCAAGATC 63148
QY 781 TACCTTATCGCGCTTCCGCTGCTCAACCTGCTCGCGTCCGAGCGCTCGAGCTGCTG 840
DB 63149 TACCTTATCGCGCTTCCGCTGCTCAACCTGCTCGCGTCCGAGCGCTCGAGCTGCTG 63208
QY 841 AAGGCGAGCGCGCGCTCATGAGAAACGCGCTGATGAGGCTGCTCGCTTCAAGATATC 900
DB 63209 AAGGCGAGCGCGCGCTCATGAGAAACGCGCTGATGAGGCTGCTCGCTTCAAGATATC 63268

```

```

QY 901 CTCAGATATGAACTGTGCTTTCGCGACGAGACCTGAGTACTGCGGGCATCGATC 960
DB 63269 CTCAGATATGAACTGTGCTTTCGCGACGAGACCTGAGTACTGCGGGCATCGATC 63328
QY 961 AAGAAAGGAGATGCTTCTCTCTGATGCCAGCGCCCTGAGATGGAATGCTATATC 1020
DB 63329 AAGAAAGGAGATGCTTCTCTCTGATGCCAGCGCCCTGAGATGGAATGCTATATC 63388
QY 1021 TCCAGCGACAGCTGTTGATGTCGACGAGGACACGCGCGAGAGCTTCCGATCGCTGCA 1080
DB 63389 TCCAGCGACAGCTGTTGATGTCGACGAGGACACGCGCGAGAGCTTCCGATCGCTGCA 63448
QY 1081 GCGCCCAATGCTGCGCGCGGGGTGTCCTGTGCTGCTCGAGCGGAGATGCGCTGAGC 1140
DB 63449 GCGCCCAATGCTGCGCGCGGGGTGTCCTGTGCTGCTCGAGCGGAGATGCGCTGAGC 63508
QY 1141 ACCATCTTCCGATGATTCGCGAGATGAAAGTGAAGAACTCCGCTTTGATATACAC 1200
DB 63509 ACCATCTTCCGATGATTCGCGAGATGAAAGTGAAGAACTCCGCTTTGATATACAC 63568
QY 1201 CCGCGCTTCCGAAATGATGATCACTCAACCTGATTTGAAGCCCTCCAAAGCTGATAG 1260
DB 63569 CCGCGCTTCCGAAATGATGATCACTCAACCTGATTTGAAGCCCTCCAAAGCTGATAG 63628

RESULT 3
AR193029 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR193029
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison,, Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Goriach,J.
TITLE Genes for the biosynthesis of epothonones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
    source
        /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match      100.0%; Score 1260; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.8e-199;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACACAGAGACGACGAAATCAGATGAGACGAAACCTGCTTCACTTCAAGCCGTTG 60
DB 62369 ATGACACAGAGACGACGAAATCAGATGAGACGAAACCTGCTTCACTTCAAGCCGTTG 62428
QY 61 GCGCCTGGGTAACGCGAGAGACCGCTTCCCGCGATGAGCGCTTGAAGAGCAACCCC 120
DB 62429 GCGCCTGGGTAACGCGAGAGACCGCTTCCCGCGATGAGCGCTTGAAGAGCAACCCC 62488
QY 121 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGCTCTGACCCGATACGACGAGTGGCG 180
DB 62489 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGCTCTGACCCGATACGACGAGTGGCG 62548
QY 181 GTGTTCCGCGACGACGCTTCCGCTGCTGATGAGAGAGTGGGAATCGACCGCGAGTAC 240
DB 62549 GTGTTCCGCGACGACGCTTCCGCTGCTGATGAGAGAGTGGGAATCGACCGCGAGTAC 62608
QY 241 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTATGTTGGGGCTGCGCGCG 300
DB 62609 TCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTATGTTGGGGCTGCGCGCG 62668
QY 301 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGCTGTTACGTACCGCGCATGAC 360
DB 62669 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGCTGTTACGTACCGCGCATGAC 62728

```

QY 361 CTGCTGCGCCGGAATACAGCGACCGTGAACAGCTGCTGATGCTGCTCGCGAGACA 420  
Db 62729 CTGCTGCGCCGGAATACAGCGACCGTGAACAGCTGCTGATGCTGCTCGCGAGACA 62788  
QY 421 GAGAGATTGACGTTGTGCGGATTAACGCGAGGGAATCCGATGCGCGATCAAGCTT 480  
Db 62789 GAGAGATTGACGTTGTGCGGATTAACGCGAGGGAATCCGATGCGCGATCAAGCTT 62848  
QY 481 CTGTTGAAGGTTCCGGCCGAGTGAACGAGAGTTCCGCTGCTTGGGCTCGCGAGCTGG 540  
Db 62849 CTGTTGAAGGTTCCGGCCGAGTGAACGAGAGTTCCGCTGCTTGGGCTCGCGAGCTGG 62908  
QY 541 CGCGGCTCGCGGTTGTTGTTGCTGCGCGAGTGAAGAGAGACCAAGACCTGCTGCG 600  
Db 62909 CGCGGCTCGCGGTTGTTGTTGCTGCGCGAGTGAAGAGAGACCAAGACCTGCTGCG 62968  
QY 601 TCCGTCACCGAGGGGCTGCGCTGCTCCATGAACGTTCTGATGAGCGGCGAGAACCG 660  
Db 62969 TCCGTCACCGAGGGGCTGCGCTGCTCCATGAACGTTCTGATGAGCGGCGAGAACCG 63028  
QY 661 CTCGAATATGACGTCTTGAACGATGCTGCTGAGCGAGGCGAGGCGAGGCTGAGC 720  
Db 63029 CTCGAATATGACGTCTTGAACGATGCTGCTGAGCGAGGCGAGGCGAGGCTGAGC 63088  
QY 721 ACGAAGAGCTGCTGCGCGCTGCTGAGGAGTATGCTGCTGAGCAGCAACATCAAGATC 780  
Db 63089 ACGAAGAGCTGCTGCGCGCTGCTGAGGAGTATGCTGCTGAGCAGCAACATCAAGATC 63148  
QY 781 TACCTTATCGGCTTGTGCTGCTCAACCTGCTGCGGCTGCGAGGCGCTGAGCTGGT 840  
Db 63149 TACCTTATCGGCTTGTGCTGCTCAACCTGCTGCGGCTGCGAGGCGCTGAGCTGGT 63208  
QY 841 AAGGCGAGCCCGGCTCATAGAGAACGCGCTGATGAGTGTCTCCGTTCCAGCAATATC 900  
Db 63209 AAGGCGAGCCCGGCTCATAGAGAACGCGCTGATGAGTGTCTCCGTTCCAGCAATATC 63268  
QY 901 CTCAGATATGAACTGTGCTGCTGCGCGAGCAGACCTGAGATCACTGCGGAGCATCATC 960  
Db 63269 CTCAGATATGAACTGTGCTGCTGCGCGAGCAGACCTGAGATCACTGCGGAGCATCATC 63328  
QY 961 AAGAAAGGAGAGTGTCTTCTCTGATCCGAGCGCTGAGAGATGGAAGCTGATTC 1020  
Db 63329 AAGAAAGGAGAGTGTCTTCTCTGATCCGAGCGCTGAGAGATGGAAGCTGATTC 63388  
QY 1021 TCCAGGCGAGAGTGTGATGAGAGCGGAGACGCGGCGAGAGCTGCGATGAG 1080  
Db 63389 TCCAGGCGAGAGTGTGATGAGAGCGGAGACGCGGCGAGAGCTGCGATGAG 63448  
QY 1081 GAGCCGCAATGCTGCGCGGAGTGTCTGCTGCTGAGAGCGGAGATGCGCGTGGG 1140  
Db 63449 GAGCCGCAATGCTGCGCGGAGTGTCTGCTGCTGAGAGCGGAGATGCGCGTGGG 63508  
QY 1141 ACCATCTTCCGAGGTTCCCGAGATGAAGCTGAAGAACTCCGTTGTTGATACAC 1200  
Db 63509 ACCATCTTCCGAGGTTCCCGAGATGAAGCTGAAGAACTCCGTTGTTGATACAC 63568  
QY 1201 CCGCGGTTCCGAGATGAATCACTCAACGTCATCTTGAAGCCTCCAAAGCTGATG 1260  
Db 63569 CCGCGGTTCCGAGATGAATCACTCAACGTCATCTTGAAGCCTCCAAAGCTGATG 63628

RESULT 4  
AR199551 68750 bp DNA linear PAT 20-APR-2002  
LOCUS AR199551 Sequence 1 from patent US 635457.  
DEFINITION AR199551  
ACCESSION AR199551.1 GI:20249625  
VERSION AR199551.1  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)

AUTHORS Schupp, T., Iigou, J., Madison, J., Molnar, I., Zirkle, R., Cyr, D., Dawn, and  
Gorlach, J.  
TITLE Genes for the biosynthesis of epoethionones  
JOURNAL Patent: US 635457-A 1 12-MAR-2002;  
FEATURES Location/Qualifiers  
SOURCE 1. 68750  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN  
Query Match 100.0%; Score 1260; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 1.8e-199; Indels 0; Gaps 0;  
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACACAGAGCAAGCAATCAAGTGAAGCAAGCTGCTTGAACCTTCAAGCCGTTTC 60  
Db 62369 ATGACACAGAGCAAGCAATCAAGTGAAGCAAGCTGCTTGAACCTTCAAGCCGTTTC 62428  
QY 61 GCGCTGGGTAACGCGAGAGACCCGTTCCGCGATGAGCGCCTGAGAGAGCAACCCCC 120  
Db 62429 GCGCTGGGTAACGCGAGAGACCCGTTCCGCGATGAGCGCCTGAGAGAGCAACCCCC 62488  
QY 121 ATCTTCTACTGAGATGAAGCCGCTCTCTGAGTCTCTCAACCCGATACACAGAGTGTGCG 180  
Db 62489 ATCTTCTACTGAGATGAAGCCGCTCTCTGAGTCTCTCAACCCGATACACAGAGTGTGCG 62548  
QY 181 GTGTTCCGCGAGCAACGCTTCCGCGTCACTGAGAGAGTGGGAAATCGAGCGCGAGTAC 240  
Db 62549 GTGTTCCGCGAGCAACGCTTCCGCGTCACTGAGAGAGTGGGAAATCGAGCGCGAGTAC 62608  
QY 241 TCGTCGCGCAATCCGAGACTGAGAGATGAAGAGTACGAGATGTTGCGGCTGCGCG 300  
Db 62609 TCGTCGCGCAATCCGAGACTGAGAGATGAAGAGTACGAGATGTTGCGGCTGCGCG 62668  
QY 301 GAGGATCAACGCTCGGCTCGCAAGCTGCTCAACCCGCTGTTACGTCACGCGCCATCGAC 360  
Db 62669 GAGGATCAACGCTCGGCTCGCAAGCTGCTCAACCCGCTGTTACGTCACGCGCCATCGAC 62728  
QY 361 CTGCTGCGCGCGGAATACAGCGACCGTGCAGCAGCTGCTGATGCTGCTCGCGAGCA 420  
Db 62729 CTGCTGCGCGCGGAATACAGCGACCGTGCAGCAGCTGCTGATGCTGCTCGCGAGCA 62788  
QY 421 GAGAGTTGCACTGTGTCGCGGATTAACGCGAGGGAATCCGATGCGCGCATCAAGCT 480  
Db 62789 GAGAGTTGCACTGTGTCGCGGATTAACGCGAGGGAATCCGATGCGCGCATCAAGCT 62848  
QY 481 CTGTTGAAGGTTCCGCGCGAGTGAAGAAATGCTGCTGCTGCGCGCATCAAGCT 540  
Db 62849 CTGTTGAAGGTTCCGCGCGAGTGAAGAAATGCTGCTGCTGCGCGCATCAAGCT 62908  
QY 541 GCGCGGCTCGCGGTTGTTGTTGCTGCGCGAGTGAAGAGAGACCAAGACCTGCTGCG 600  
Db 62909 GCGCGGCTCGCGGTTGTTGTTGCTGCGCGAGTGAAGAGAGACCAAGACCTGCTGCG 62968  
QY 601 TCCGTCACCGAGGGGCTGCGCTGCTCCATGAACGTTCTGATGAGCGGCGAGAACCCG 660  
Db 62969 TCCGTCACCGAGGGGCTGCGCTGCTCCATGAACGTTCTGATGAGCGGCGAGAACCCG 63028  
QY 661 CTCGAATATGACGTCTTGAACGATGCTGCTGAGCGAGGCGCGAGCGAGCTGAGC 720  
Db 63029 CTCGAATATGACGTCTTGAACGATGCTGCTGAGCGAGGCGCGAGCGAGCTGAGC 63088  
QY 721 ACGAAGAGCTGCTGCGCGCTGCTGAGGAGTATGCTGCTGAGCGAGATACACGATC 780  
Db 63089 ACGAAGAGCTGCTGCGCGCTGCTGAGGAGTATGCTGCTGAGCGAGATACACGATC 63148  
QY 781 TACCTTATCGGCTTGTGCTGCTCAACCTGCTGAGGAGTGTGCGCGAGCTGAGCTGGT 840  
Db 63149 TACCTTATCGGCTTGTGCTGCTCAACCTGCTGAGGAGTGTGCGCGAGCTGAGCTGGT 63208  
QY 841 AAGGCGAGCCCGGCTCATAGAGAACGCGCTGATGAGTGTCTCGCTTCCAGCAATATC 900  
Db 63209 AAGGCGAGCCCGGCTCATAGAGAACGCGCTGATGAGTGTCTCGCTTCCAGCAATATC 63268

QY 901 CTCAGAAATGAACTGTGCTTCCGACGACGACCTGAGTACTCCGGGCGATGATC 960  
DB 63269 CTCAGAAATGAACTGTGCTTCCGACGACGACCTGAGTACTCCGGGCGATGATC 63328  
QY 961 AAGAAAGGAGATGATGCTTCTTCTCTGATCCGAGGCGCTGAGATGGAGCTGATTC 1020  
DB 63329 AAGAAAGGAGATGATGCTTCTTCTCTGATCCGAGGCGCTGAGATGGAGCTGATTC 63388  
QY 1021 TCCAGGCCAGACGTGTTGATGTCAGACGAGGCGGACGCTCGTACGCTGAG 1080  
DB 63389 TCCAGGCCAGACGTGTTGATGTCAGACGAGGCGGACGCTCGTACGCTGAG 63448  
QY 1081 GGGCCCATGTCGCCCCGGGGGTGTCCTTCTGCTGCTGAGGCGGAGATCCGCTGGG 1140  
DB 63449 GGGCCCATGTCGCCCCGGGGGTGTCCTTCTGCTGCTGAGGCGGAGATCCGCTGGG 63508  
QY 1141 ACCATCTTCGTAAGTTCCTCCGAGATGAAGCTGAAGAACTCCGTTTGGATACAC 1200  
DB 63509 ACCATCTTCGTAAGTTCCTCCGAGATGAAGCTGAAGAACTCCGTTTGGATACAC 63568  
QY 1201 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 1260  
DB 63569 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 63628

RESULT 5  
ARI99559  
LOCUS ARI99559 68750 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6355458.  
ACCESSION ARI99559  
VERSION ARI99559.1 GI:20249633  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)  
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.  
TITLE Genes for the biosynthesis of epothilones  
JOURNAL Patent: US 6355458-A 12-MAR-2002;  
FEATURES  
LOCATION/Qualifiers  
source 1. 68750  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN

Query Match 100.0%; Score 1260; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 1.8e-199;  
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATGACACAGAGACGAAATCAGATGAGACGAGCTGCTTCACTTCAAGCCGTTTC 60  
DB 62369 AATGACACAGAGACGAAATCAGATGAGACGAGCTGCTTCACTTCAAGCCGTTTC 62428  
QY 61 GGGCCCTGGGTAACCGGAGACCCGTTTCCCGCATGAGGCGCTTGAAGAGGAAACCCC 120  
DB 62429 GGGCCCTGGGTAACCGGAGACCCGTTTCCCGCATGAGGCGCTTGAAGAGGAAACCCC 62488  
QY 121 ATCTTCTACGTGAGTAAGGCGCTCTGGGTCCTCAACCGATACACGAGCTGTGGG 180  
DB 62489 ATCTTCTACGTGAGTAAGGCGCTCTGGGTCCTCAACCGATACACGAGCTGTGGG 62548  
QY 181 GTGTTCCGCGACGACGCTTCCGCGTCACTGAGAAAGTGGAAATCGAGCGGAGTAC 240  
DB 62549 GTGTTCCGCGACGACGCTTCCGCGTCACTGAGAAAGTGGAAATCGAGCGGAGTAC 62608  
QY 241 TCGTCGGCACTTCCGAGCTCAGACGATATGAAGAATGAGATGTTTCCGGCTGGCCG 300  
DB 62609 TCGTCGGCACTTCCGAGCTCAGACGATATGAAGAATGAGATGTTTCCGGCTGGCCG 62668  
QY 301 GAGGATACGCTCGGGTCCGCAAGCTGTCACCCCGTCTTACGTCACGCGCATGAC 360

DB 62669 GAGGATACGCTCGGGTCCGCAAGCTGTCACCCCGTCTTACGTCACGCGCATGAC 62728  
QY 361 CTGCTGCGGCGCGAAATACAGGCAACCGTCCAGACGCTGCTGATGCTGCTCCGACAA 420  
DB 62729 CTGCTGCGGCGCGAAATACAGGCAACCGTCCAGACGCTGCTGATGCTGCTCCGACAA 62788  
QY 421 GAGGATGCAAGCTTGTGCGGATTAACGCGAGGAGATCCGATGCGCGCATGACGCT 480  
DB 62789 GAGGATGCAAGCTTGTGCGGATTAACGCGAGGAGATCCGATGCGCGCATGACGCT 62848  
QY 481 CTGTTGAAGGTTCCGCGCAGATGTCAGAGAGTTCCGTCGCTTGGGCTCGGCGACTG 540  
DB 62849 CTGTTGAAGGTTCCGCGCAGATGTCAGAGAGTTCCGTCGCTTGGGCTCGGCGACTG 62908  
QY 541 GCGCGCTCGCGGTGTTGGTGGCCCAAGTTCATGAGAGAACCAAGACCTGGTCCG 600  
DB 62909 GCGCGCTCGCGGTGTTGGTGGCCCAAGTTCATGAGAGAACCAAGACCTGGTCCG 62968  
QY 601 TCCGTCACCGAGGGGCTGGCGCTGCTCATGACGTCCTCATGAGCGGGCGAGAACCG 660  
DB 62969 TCCGTCACCGAGGGGCTGGCGCTGCTCATGACGTCCTCATGAGCGGGCGAGAACCG 63028  
QY 661 CTCGAAAATGACGTCCTTGAACATGCTGCTTCAAGCCGAGCCGACGAGCGCTGAGC 720  
DB 63029 CTCGAAAATGACGTCCTTGAACATGCTGCTTCAAGCCGAGCCGACGAGCGCTGAGC 63088  
QY 721 ACGAAGACGTGTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
DB 63089 ACGAAGACGTGTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63148  
QY 781 TACCTTATCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
DB 63149 TACCTTATCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63208  
QY 841 AAGCCGAGCCCGGCTCATGAGAACGCGCTCATGAGTGTGCTGCTGCTGCTGCTGCTGCT 900  
DB 63209 AAGCCGAGCCCGGCTCATGAGAACGCGCTCATGAGTGTGCTGCTGCTGCTGCTGCTGCT 63268  
QY 901 CTCAGAAATGAACTGTGCTTCCGACGACGACCTGAGTACTCCGGGCGATGATC 960  
DB 63269 CTCAGAAATGAACTGTGCTTCCGACGACGACCTGAGTACTCCGGGCGATGATC 63328  
QY 961 AAGAAAGGAGATGATGCTTCTTCTCTGATCCGAGGCGCTGAGATGGAGCTGATTC 1020  
DB 63329 AAGAAAGGAGATGATGCTTCTTCTCTGATCCGAGGCGCTGAGATGGAGCTGATTC 63388  
QY 1021 TCCAGGCCAGACGTGTTGATGTCAGACGAGGCGGACGCTCGTACGCTGAG 1080  
DB 63389 TCCAGGCCAGACGTGTTGATGTCAGACGAGGCGGACGCTCGTACGCTGAG 63448  
QY 1081 GGGCCCATGTCGCCCCGGGGGTGTCCTTCTGCTGCTGAGGCGGAGATCCGCTGGG 1140  
DB 63449 GGGCCCATGTCGCCCCGGGGGTGTCCTTCTGCTGCTGAGGCGGAGATCCGCTGGG 63508  
QY 1141 ACCATCTTCGTAAGTTCCTCCGAGATGAAGCTGAAGAACTCCGTTTGGATACAC 1200  
DB 63509 ACCATCTTCGTAAGTTCCTCCGAGATGAAGCTGAAGAACTCCGTTTGGATACAC 63568  
QY 1201 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 1260  
DB 63569 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 63628

RESULT 6  
ARI99567  
LOCUS ARI99567 68750 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6355459.  
ACCESSION ARI99567  
VERSION ARI99567.1 GI:20249641  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)  
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.  
TITLE Genes for the biosynthesis of epothilones  
JOURNAL Patent: US 6355459-A 12-MAR-2002;  
FEATURES  
LOCATION/Qualifiers  
source 1. 68750  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN





Db	62669	GAGGATCACGCTCGGGTCCCGAGCTGTCAACCCGCTGTTTAAGTCACGGCGCATCGAC	62728
OY	361	CTGCTGCGGCGCCGAATATACAGCGCACCTGTGCACCAAGCTGCTCGATGCTGCTCCGGACAA	420
Db	62729	CTGCTGCGGCGCGCAAAATACAGCGCACCGGTGCACCAAGCTGCTCGATGCTGCTCCGGACAA	62788
OY	421	GAGGAGTTCCGACGTTTGCGGGGAAATTACGGGAGGGAAATCCCAATGGCGCGATCAGCGCT	480
Db	62789	GAGGAGTTCCGACGTTTGCGGGGAAATTACGGGAGGGAAATCCCAATGGCGCGCGATCAGCGCT	62848
OY	481	CTGTTGAAGGTTCCGGCGCAAGTGTGACGAAAGTTCGTTGCTGCTCGGCTCGGCACTGCG	540
Db	62849	CTGTTGAAGGTTCCGGCGCAAGTGTGACGAAAGTTCGTTGCTGCTCGGCTCGGCACTGCG	62908
OY	541	CGCGCGCTCGGCGGTGGGTTTGTTGTCGCCCAAGTCGATGAGGAGCCAAAGCCTGTGTGCG	600
Db	62909	CGCGCGCTCGGCGGTGGGTTTGTTGTCGCCCAAGTCGATGAGGAGCCAAAGCCTGTGTGCG	62968
OY	601	TCCGTCACCCAGGGGGGCTCGCGCTGCTGCATGACGTCCTCGATAGGGGGCGCAGAGAACCG	660
Db	62969	TCCGTCACCCAGGGGGGCTCGCGCTGCTGCATGACGTCCTCGATAGGGGGCGCAGAGAACCG	63028
OY	661	CTCGAAAATGACGTCCTTGAAGATGCTGCTTCAAGGCGAGGCGCAGAGGACAGAGCTGAGC	720
Db	63029	CTCGAAAATGACGTCCTTGAAGATGCTGCTTCAAGGCGAGGCGCAGAGGACAGAGCTGAGC	63088
OY	721	ACGAGGAGCTGATGCGCGCTCTGTGGGTGCGAATATATGCTGTGCTGCGACCGATACCAAGATC	780
Db	63089	ACGAGGAGCTGATGCGCGCTCTGTGGGTGCGAATATATGCTGTGCTGCGACCGATACCAAGATC	63148
OY	781	TACCTTATCGGGTTCGCTGTGTCTCAACCTGTGTGGGTGCGCGCGAGGCGCTCCAGCTGGTG	840
Db	63149	TACCTTATCGGGTTCGCTGTGTCTCAACCTGTGTGGGTGCGCGCGAGGCGCTCCAGCTGGTG	63208
OY	841	AAGGCGGAGCCCGGGCTCATAGAGGAAGCGCGCTCGATAGAGGTGCTCGGCTTCGACAAATATC	900
Db	63209	AAGGCGGAGCCCGGGCTCATAGAGGAAGCGCGCTCGATAGAGGTGCTCGGCTTCGACAAATATC	63268
OY	901	CTCAGAAATAGGAACCTGTGCGTTTGCCAGGACAGGACCTGTGAGTACTGCGGGGATCGATC	960
Db	63269	CTCAGAAATAGGAACCTGTGCGTTTGCCAGGACAGGACCTGTGAGTACTGCGGGGATCGATC	63328
OY	961	AAGGAAGGGGAGATGATCTTTCTCTTATATCCCGAGGCGCTTGAGAGATGGGACTGTATTC	1020
Db	63329	AAGGAAGGGGAGATGATCTTTCTCTTATATCCCGAGGCGCTTGAGAGATGGGACTGTATTC	63388
OY	1021	TTCAGGGCCAGACGATTTGATGTGCGAGGGGACACGGGCGCGAGCTCGCGATACGGGTAGA	1080
Db	63389	TTCAGGGCCAGACGATTTGATGTGCGAGGGGACACGGGCGCGAGCTCGCGATACGGGTAGA	63448
OY	1081	GGCCGCCCATGTCTGCCCCCGGGGTGTCCCTTGCTGCTCGAGCGGAGAGATCGCCGTGGGC	1140
Db	63449	GGCCGCCCATGTCTGCCCCCGGGGTGTCCCTTGCTGCTCGAGCGGAGAGATCGCCGTGGGC	63508
OY	1141	ACCATCTTTCGTAGGTTCCCGAGATGAAGCTGAAGAAATCTCCGTGTTTGATATCCAC	1200
Db	63509	ACCATCTTTCGTAGGTTCCCGAGATGAAGCTGAAGAAATCTCCGTGTTTGATATCCAC	63568
OY	1201	CCCGGTTTCGGGAACATCGAATCACTCAAGGTATCTGAAGGCGCTCCAAAGCTGGATAG	1260
Db	63569	CCCGGTTTCGGGAACATCGAATCACTCAAGGTATCTGAAGGCGCTCCAAAGCTGGATAG	63628

[illegible]

REFERENCE	1 (bases 1 to 68750)	Unclassified.
AUTHORS	Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and Goriach,J.	
TITLE	Genes for the biosynthesis of epothilones	
JOURNAL	Patent: US 6383787-A 1 07-MAY-2002;	
FEATURES	Location/Qualifiers	
source	1..68750	
BASE COUNT	9596 a 22456 c 25539 g 11159 t	/organism="unknown"
ORIGIN		
Query Match	100.0%; Score 1260; DB 6; Length 68750;	
Best Local Similarity	100.0%; Pred. No.1,8e-199;	
Matches 1260;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGACACAGGAGCAGATCGATGAGACGAACTGCTTTCGACTTCAAGCCGTTTC	60
DB	62369 ATGACACAGGAGCAGATCGATGAGACGAACTGCTTTCGACTTTCGACTTTCGACTTTC	62428
QY	61 GCCCTGGGATACCCGAGGACCCGTTTCCCCCGATGAGACGCGCTGAGAGGCAACCCCC	120
DB	62429 GCCCTGGGATACCCGAGGACCCGTTTCCCCCGATGAGACGCGCTGAGAGGCAACCCCC	62488
QY	121 ATCTTCTACTGGGATGATGAGGCGCGCTCTGAGGCTCTCAACCGATACACGACGCTGCGCG	180
DB	62489 ATCTTCTACTGGGATGATGAGGCGCGCTCTGAGGCTCTCAACCGATACACGACGCTGCGCG	62548
QY	181 GTGTTCCGCGACGACGCTTTCGCGGTCAGTCGAGAGAGTGGGAAATCGACGCGGAGTAC	240
DB	62549 GTGTTCCGCGACGACGCTTTCGCGGTCAGTCGAGAGAGTGGGAAATCGACGCGGAGTAC	62608
QY	241 TCGTCGCGCATTTCCGAGCTCAGCGATATGAGAAAGTACGATTTGTTGGGCTGCGCGCG	300
DB	62609 TCGTCGCGCATTTCCGAGCTCAGCGATATGAGAAAGTACGATTTGTTGGGCTGCGCGCG	62668
QY	301 GAGGATACGCTGGGGTCGCGAAGCTGCTCAACCGCTGTTTACGTACGCGCATTCGAC	360
DB	62669 GAGGATACGCTGGGGTCGCGAAGCTGCTCAACCGCTGTTTACGTACGCGCATTCGAC	62728
QY	361 CTGCTGCGCGCCGAAATATACGCGCACCGCTGACCAAGCTCTCGATGCTTCCTCCGACAA	420
DB	62729 CTGCTGCGCGCCGAAATATACGCGCACCGCTGACCAAGCTCTCGATGCTTCCTCCGACAA	62788
QY	421 GAGGATTCGACGTTGTGCGGGATTACGCGGAGGAGATCCGATGCGCGCATCAGCGCT	480
DB	62789 GAGGATTCGACGTTGTGCGGGATTACGCGGAGGAGATCCGATGCGCGCATCAGCGCT	62848
QY	481 CTGTGAAGGTTCCGGCCGACGATGTGACGAAAGTTCCGTGCGTGGCGCTCGGCGACTGCG	540
DB	62849 CTGTGAAGGTTCCGGCCGACGATGTGACGAAAGTTCCGTGCGTGGCGCTCGGCGACTGCG	62908
QY	541 CGCGCGCTCGCGCTGGGTTTGGTGCCCGCAGCTGATGAGAGACCAAGACCTGTGTGCG	600
DB	62909 CGCGCGCTCGCGCTGGGTTTGGTGCCCGCAGCTGATGAGAGACCAAGACCTGTGTGCG	62968
QY	601 TCCGTCACGAGGGGCTCGCGCTGCTCATGACGCTCTCGATGAGGCGCGCAGAAACCG	660
DB	62969 TCCGTCACGAGGGGCTCGCGCTGCTCATGACGCTCTCGATGAGGCGCGCAGAAACCG	63028
QY	661 CTGGAATAATGCTCTTTCGACGATGCTGCTTACGCGCCGAGCGGAGCGGCGAGGCTGAGC	720
DB	63029 CTGGAATAATGCTCTTTCGACGATGCTGCTTACGCGCCGAGCGGAGCGGCGAGGCTGAGC	63088
QY	721 ACGAAGAGGTGGTCGCGCTCGTGGGTGCGATATGCTGCTGTGGCACCGATACAGATC	780
DB	63089 ACGAAGAGGTGGTCGCGCTCGTGGGTGCGATATGCTGCTGTGGCACCGATACAGATC	63148
QY	781 TACCTTATCGCTTTCGCTGTGCTCAACTGCTCGGCTCGCCGAGCGGCTGTGAGCTGTG	840
DB	63149 TACCTTATCGCTTTCGCTGTGCTCAACTGCTCGGCTCGCCGAGCGGCTGTGAGCTGTG	63208
QY	841 AAGCGCAGCCCGGCTCATAGAGAAACGCGCTGATGAGAGTGTCCGCTTCGACAAATATC	900

Df	63209	AAGCCGACGCCGGCTCATATAGAGAACCGCCTGATAGGTGCTCCGCTTCACAATATTC	
OY	901	CTCAGAATTAGGAACCTGTCGCTTTGCACCAGGACAGAACTGGAGTACTCGCGGACATTCATC	960
Df	63269	CTCAGATATGGAAGTAAGTGCGCTTTTCGCCAGGACAGAACTGGAGTACTCGCGGACATTCATC	63328
OY	961	AAGAAAAGGAGATGATGCTTTCTCTCTGATCCGACGCGCTTGAGATGGACTGTATTC	1020
Df	63329	AAGAAAAGGAGATGATGCTTTCTCTCTGATCCGACGCGCTTGAGATGGACTGTATTC	63388
OY	1021	TCCAGGCGACACGCTTTGATATGTCGACGGGACACCGGCGCGCTCCGCTGATACGCTATA	1080
Df	63389	TCCAGGCGACACGCTTTGATATGTCGACGGGACACCGGCGCGCTCCGCTGATACGCTATA	63448
OY	1081	GCCCCCATGCTGCGCCCCGGGGGTGCTCCCTTGCTCGCTCGAGGCGAGATTCGCCGTGGC	1140
Df	63449	GCCCCCATGCTGCGCCCCGGGGGTGCTCCCTTGCTCGCTCGAGGCGAGATTCGCCGTGGC	63508
OY	1141	ACCATCTTCGCTAGTGTCCCGAGATGAAGCTGAAAAAATCCCGCTGTTGGATACAC	1200
Df	63509	ACCATCTTCGCTAGTGTCCCGAGATGAAGCTGAAAAAATCCCGCTGTTGGATACAC	63568
OY	1201	CCCCCGTTCCGGAAATCGAATCACTCAAGCTATCTTGAAGCGCTTCGAAAGCTGATATG	1260
Df	63569	CCCCCGTTCCGGAAATCGAATCACTCAAGCTATCTTGAAGCGCTTCGAAAGCTGATATG	63628
RESULT_9			
LOCUS	AF217189	58733 bp	DNA linear BCT 09-JUN-2000
DEFINITION	Sorangium celulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.		
VERSION	AF217189		
KEYWORDS	AF217189.1 GI:7453554		
SOURCE ORGANISM	Polyangium celulosum Polyangium celulosum Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium. 1 (bases 1 to 58733) Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B. Cloning and heterologous expression of the epothilone gene cluster Science 287 (5453), 640-642 (2000)		
REFERENCE	JOURNAL MEDLINE PUBMED		
AUTHORS	Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L. and Khosla, C. Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium celulosum Gene 249 (1-2), 153-160 (2000)		
TITLE	3 (bases 1 to 58733) Julien, B. Direct Submission Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA Location/Qualifiers 1..58733 /organism="Polyangium celulosum" /mol_type="genomic DNA" /strain="SMP4" /db_xref="taxon:56" <1..992 /note="orf4" /codon_start=3 /trans_table=11 /product="putative transposase" /protein_id="AAF62878.1"		
JOURNAL MEDLINE PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
CDS			

```

/d_b_xref="GI:7453555"
/translation="VRGVEAFNDFPGVGVLYLNDLNLKAVNRHGDAIRFPHITLAL
SADYFEERPEPAVARGNKEGFVERAIRIVVEGFPEAKAVADLIDLNRQATIEWTSAL
DRSTRELLAVTRQAFDFDERSVNLIRHDTPEPHIERVEVGGTTPVAPRQPLNAPVH
DRTRETVLIVLADLSOVLADNOJLVAHVNSMDGQOIQEPHQLQVLDEKRAERER
GLDRLAPASRSCQAFIRIVARGNGVSSAIALRLQILIDAGALEEALVEVLEVR7I
HIGAVRQYIDIRRSERHLPPEVSIPTVRGEHALVTTPHSLTTYDALKQDPT"
989..1501
/feature="orfB"
/codon_start=1
/translation="putative transposase"
/protein_id="AAFG2879.1"
/db_xref="GI:7453556"
/translation="MTDLTPETRKLSLGLFGLLACWEQLADKPELVREVLAREE
RHKSIERLKNRVAAFKPMTDSDMPKKIRREAVDDLYDSRYADLLFEVTRRD
AOKPLISTNKAPAFMDQGVFPHACVTVLVDLVRHRAVEIEAEASRYLKEALNAT
RTKQRTKKH"
1998..6263
/gene="epob"
1998..6263
/gene="epob"
/gene="epob"
/feature="type I polyketide synthase; epothilone PKS loading
module"
/codon_start=1
/translation="EpoA"
/product="EpoA"
/protein_id="AAFG2880.1"
/db_xref="GI:7453557"
/translation="MADRPIERAAEDPIALVGNACRLPGVINDLSGFWTLDESSRTV
QGVPERMDAAMPDDLDGKTPVTRASFDSVACFDASFGTISRELRNDPAPR
LLEVERKALNNAAIAPBALGTEGTGVIIGIPEPEXALPRATASKEIDAHGIGCTM
PSVGRASIVYLIGRGPCVAVDITASLSLVAVHLACSLGEGESTLACGVSILMSP
STVGLASTRYLATDGRCAFSADGDFGEGCAVVVLIRSGARADGRILAVIRG
SAINDDGASGLTYPNGSSOEIVKRLADAGCAASGVGVHAGSTTIGDPIEIOA
LNAVYGLRDVATPULGIYKYNLGRHEVYASSTIGLLKYVLSLQHQI PAHLAQLN
PRISGDLRLVTARPTWPMWTPRRAGVSSFGMSGTNAHVLEAPATCTPRAE
PRABELLYASATAALDAHARLDRHLETPSOQLGVAASLITRSAMERLAVAT
SEEGRALDASQAQGPVGVRR1IADSSGKLAFLTECGAQTLLGWRGLYVWAF
REAPFLCYRLNOELDRPLREAVGAFSEADVADLADOTAPALFPEEVALAAMS
WCEPFLVAGHSIGIEVAACVAGVSEADVAFVLAAGTRMQLPGGMAVLAABA
DYAAVAVHAASVSIAAVNGPDQVYIAGACOPHATLAAAMAAGARTKALHSHAHFS
PLMAAMEAFGRVAVSVSYRRPSIVLNSLNSGKAGDVESSPGVWHAHVVRP
AVGGLVSNAGLEPGSGRRVLPPTYPMQERWYIDTKADDAAGDRAPAGHKEVEKG
GAVRGDRSRRLDHPPESSGRREKVEVAGADRPRELIDBPGLDRLVLRTERBAAG
LGEVYIANDAGLSFNDYQALAKMVPDDLGEKNPPLLIGEGECAGRI VAGSEVNLV
VGQPIVLSAQAFAHVHTTSALVLRPQALSTEAAMVATLTMATLADSGTARQP
GERVLIHAATGGVGLAVQMAOHGAVHATAGIPEKRAVLESLGVARYVDSRSDIP
ADVRSMTGDEGVADVNLISGELIDKSFNLIRSGRVELEKGDVCVANDOLGIRPTLR
NLSEFLVLRGMLERPARVARVLEELIGLIAGVFPPTATLPTARVADARSMQO
AOKHDKVLTLDGPBEVOR1PTTHAGAGPTGDBDLDRLSAKAPABAALAEAR1TO
VSQVLRTEIYVAGKAEFTRLGMDSLMAVELRNRIETSLKSLTTPFLSTSNIALIT
ONLUDALATSLERVAENLRAGVSDPVSSGADQDMEITALL"
6260..10492
/gene="epob"
6260..10492
/gene="epob"
/feature="non ribosomal peptide synthetase; epothilone PKS
module 1"
/codon_start=1
/translation="EpoB"
/product="EpoB"
/protein_id="AAFG2881.1"
/db_xref="GI:7453558"
/translation="MTINQLNLBLEHGQVKLADGERLQIQAPKALNPMLARISEH
KSTIDVLTQRQLPAESIVPAEABRRPPLPTDQGSWLGRTAFAFVPSLIRNAYRED
CTDLDVALTSRQAPKRVVAKHMLRAHTLPMWQVIEKVDADIEIDLRLDSTSEYD
RLVLSRDMNSHRIYDTERPPLXYHVAARLDEOQRLVLSIDLINVLGSLSIIFKML
SEYDEPESLIVLEISYRDVYLAESRKSBAHORSMDWKRRVALLPPPMKAD
PSYLAIEIFRTEOMLPDSDWSRLKQVGERGULTPTGVYIILASVYGRKASPRTLN
GTEVREARAPVIRGLDGETSMVLIDITTRTKSEGRKAKIIOEVLAMFAMCHVS
NILEVREARAPVIRGLDGETSMVLIDITTRTKSEGRKAKIIOEVLAMFAMCHVS
NILEVREARAPVIRGLDGETSMVLIDITTRTKSEGRKAKIIOEVLAMFAMCHVS

```

```
0. YEHDDGLVLA MDIVDGVPPDLDLMDLCAVVAFLRLTEBPMGSRMCSP, PAOLE
AASANETMSLSLSEHTLHGLFAARVEQLEPMQALAVSARKTITYEBELSRSRRLGARLR
EGGARNPTLVA VVMEKGEQVAVVAVLAVESGA VYFIDDLPAERLHYLLDGEVXLV
LQPMWLDGLSWPBGIOQLVSDAVGEGDQLPMWVLOTPSDLAVVYITSTDEYLV
GVMIDHRCALVNTLIDINERFEIGPGRVATLSLSLSDLSVYDFGLAAGTIVVPA
SKLRDPAHMAALIEREKTVVMSVPALMELVEHSEGRDLSARSRLSLSDGMDV
GLBGLQAIRPGCVSVISLIGATEASISWISGYPVARNPSMASIPYGRPLRNTQTHYLD
BALBRRPVPPGQVITGCVGLAGTIRDEKTRNSFLVHPTEGERLXKTDGDLGRVLP
GNI EFMGREDNOIKLKGIRVEGEIETLSKSHPNVADAVIVGANDAKALLAVVP
EOTRRRAAQDQSLKTERVDARAAHAKADGLDGERVQKLAHGLRLDLKDPVVDL
TGLVPEAGLDVYARRSVRTLEAVIPVEFERFLSCLSVEPDGALAKRRYSAG
STYPVOTVYAKSGRLEGVDEGFPYHPEHHLKMDSDIERGAVPWPNPVPEBA
PULLVGRIDATLESYLSREBCELEAGYMLLMEQAPSCNIGCPGQDPEOVR
PYLDRHSVYVYHGLMGRVDRQVCTLGDSSRRRTTGAPRGROHRAIDLID
FRITKLPEYVFPVFEVLDALPUTSNGKDRALKERKOTSSPRSHGHTAPDABEII
LVA VVEVLGLEVGLQOSFVLDGATSIHVMSRSLQRLDRELAITELFOYPMIGS
LASGLRDSKDLDEORPMQDRVEARKRRRS"
10639..16137
/gene="epoc"
10639..16137
/note="type I polyketide synthase; epochlione PKS module
2"
/codon_start=1
/cranel_table=11
/product="Bpoc"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEQESSAIAVIGMSGRFGARLDLDFMRNLJDTCEAVQRESBO
BLASGVDPALVLDPMYVAGSVLEEDVDFDAFPISSREALMDPORIMECME
ALEMNGYDPTAYEGSIGVYAGANMSYLTLSNHEHPANMGITPTILIGNKDIAT
HYSRLNLGSPISVOTACSTSLVAVHLACMSLHRECMALAGITVIVIPRAGVY
AAGGILSPDGHCAFPKAKNGITTMGNCVVLKLPDLRLSGDPRAVILISATNDV
GAKIGFAPSEVGQAQALMEALALAGVARSIOYIETHTGTLLDLETALETRVFE
DADASTRSCAIGSVKTIQGLHESAGVAGSLIKTYVLEHRTQLPSLNESSPNISDE
ASSPFYNTSLKDMWTGSTPRAGVSSFGIGGTNAVIVLEAPAAKLPAAPARAEAL
FVYSAKSAALADAAARLDHLOAHQGLSDGVAFSLATVSEMERLMAAPSEBAL
REGDAARAGOTPGAVRGRCSPGVPRVYVFPPOGSGWGMQOLLAERVFHAL
SACDRAIQBAKMSLAEALADESSQIETRIDVQVYLPALAVPRLMRSNGEDV
VIGHSGEVAHAHVAGALSLEDAVALICRSRLRRI SQGEMAVTELSLAEAGLR
GYEDRVSVA VNSPRSTVLSGEPALIGEVLSSINAGVCRVKVDVASHSPQVPLR
EDLLAALGILRPRAAVPMRSTVTGAMVGPPELGANVMNNIRQVPRFAEVQAOGLQ
GHLGFEVMSPHPLITTSVEEMRAAQRBAAGVLSRGDSEPMLEALGALGALGOST
VWGRLFPAGGRBRLPTYPMQREPVYITAPAKSAAGDRGPAHGHPLGMMQITST
QYSTRIMETTLDLKLPMLGDRVQCAVVPFPAAYIEMALISSAGEALGCPYQITDV
LAELAFAGDAAVLVQVVTTEQPSGRLOQIASRAGACHASFRVHARGALLRVERTE
VPAGLTLASVARRLQASMPAAATYAEITLMEGLQYGPAPFGIIELMNGEBELGRVLP
DAAGSAAEYRLPALDADCFOVVGLFAGGGEATPMVPEVGSRLLOPSELMECHA
RVNHGROTPDRQGDPMVTVSSGA VAVASGLVQRLPGVRREDDEWPLFLEMEP
AAVGTAKVNAAGRMILGGGGGGAALRSWLEAGHAYVHAASNTSAACTVRLAKAF
DCQAPPAVYHLSLDGGGELDPQLGAGHLDAPRADVSPDALDPAVLCDSVLMYV
QALAGMFRDAPRLWLITRGAQAVGADVSTQAPLIGRVIYAMHADURCARVLDL
PRPDELGALLAEALADDAEVALRGERCVARIVRQOPETPRGRJESCVPTDVT
IADSTYLVLTGGIGLGSVAGLAEGRGAGHLVLRSGAASVEQAAVAALAEAGAR
VTVAKADVADRAOLERILREVTSGMVLGVVAAAGIINDGLMOOTPRPFXMAPK
VOGALHALLTBEARLSPFVYVAGVGLGSPQCGYAAATGFLMLAHRRXAOGLPA
LEVDWGLPAVEGMAAQBDERGARLSRGRKSLTPDGGLSALAKRLESQVQGNPVN
PRUWELYPAAASSRMLSLVTRHRSAGPAGDGLRLHGAETSAASGJLEBLRL
AOISOYLRLEBGEIENDAPLTLGNMSLGELEIRNLEIMLGIETVATLMTWYPTVAA
LEGLHAREACBAAPVSPHTTADSAVEIEEMSDDLTOLIAAKFKALT"
16134..37907
/gene="epod"
16134..37907
/note="epod"
/gene="epod"
/note="type I polyketide synthase; epochlione PKS modules
3, 4, 5, and 6"
/codon_start=1
/cranel_table=11
/product="epod"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAOQNPLKOAIIIORELSRAGLAQALERTPEPIAV
GIGCRPGADAPBEATFELLDAERDAVQPLDRRMALVGAPVAEAVPHMAGLTPEDIC
```

Db 57597 AAGCGGAGCCGGGCTCATGAGGAACGGCTCATGAGTCTCGCTTCACAAATATC 57656  
Qy 901 CTCAGATATGGAACGTGTGCTTTCGACAGGAGGAGCCTGAGTACTGGGGGAGCATCATC 960  
Db 57657 CTCAGATATGGAACGTGTGCTTTCGACAGGAGGAGCCTGAGTACTGGGGGAGCATCATC 57716  
Qy 961 AAGAAAGGAGAGTGTCTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 1020  
Db 57717 AAGAAAGGAGAGTGTCTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 57776  
Qy 1021 TCCAGGCCAGACGTGTGTGATGTGACCGGAGACAGGCGCGAGCCTGCGGTAGAC 1080  
Db 57777 TCCAGGCCAGACGTGTGTGATGTGACCGGAGACAGGCGCGAGCCTGCGGTAGAC 57836  
Qy 1081 GGGCCCAATGTCTGCCCCGGGGGTGTCCTGCTGCTGCTGAGGCGGAGATGCGCGTGGGC 1140  
Db 57837 GGGCCCAATGTCTGCCCCGGGGGTGTCCTGCTGCTGCTGAGGCGGAGATGCGCGTGGGC 57896  
Qy 1141 ACCATCTTCCGTAGGTCCCGAGATGAGCTGAAAGAACTCCGCTGTTTGATATCAC 1200  
Db 57897 ACCATCTTCCGTAGGTCCCGAGATGAGCTGAAAGAACTCCGCTGTTTGATATCAC 57956  
Qy 1201 CCGCGTTCCGGAACATGATCACTCAACGTCACTTTGAGACCTCCAAAGCTGATTA 1259  
Db 57957 CCGCGTTCCGGAACATGATCACTCAACGTCACTTTGAGACCTCCAAAGCTGATTA 58015

RESULT 10  
ARI72664  
LOCUS ARI72664 71989 bp DNA linear PAT 17-DEC-2001  
DEFINITION Sequence 2 from patent US 6303342.  
ACCESSION ARI72664  
VERSION ARI72664.1 GI:17912155  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 71989)  
AUTHORS Julien B., Katz L., Khosla C. and Tang L.  
TITLE Recombinant methods and materials for producing epothenones C and D  
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;  
FEATURES  
source 1..71989  
/organism="unknown"  
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others  
ORIGIN

Query Match 98.9%; Score 1246.2; DB 6; Length 71989;  
Best Local Similarity 99.4%; Pred. No. 3.4e-197;  
Matches 1251; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGACACAGAGCAAGCCGATCAGAGTGAAGACGAGGCTGCTTTCGACTTCAAGCCGTTTC 60  
Db 56757 ATGACACAGAGCAAGCCGATCAGAGTGAAGACGAGGCTGCTTTCGACTTCAAGCCGTTTC 56816  
Qy 61 GCGCGTGGGTACGCGGAGGAGCCGTTCCCGCGATCGAGCGCTGAGAGGAGCAACCCCC 120  
Db 56817 GCGCGTGGGTACGCGGAGGAGCCGTTCCCGCGATCGAGCGCTGAGAGGAGCAACCCCC 56876  
Qy 121 ATCTTCTACTGAGTGAAGGCGCTCTTGGTCTCTACCCCGATACACAGCTGTGCGG 180  
Db 56877 ATCTTCTACTGAGTGAAGGCGCTCTTGGTCTCTACCCCGATACACAGCTGTGCGG 56936  
Qy 181 GTGTTCGCGAGCAACGCTTCCGCGTCACTGAGAGAGTGGGAATCGAGCCGAGATAC 240  
Db 56937 GTGTTCGCGAGCAACGCTTCCGCGTCACTGAGAGAGTGGGAATCGAGCCGAGATAC 56996  
Qy 241 TCGTGGGCGCATTCGCGAGCTGACGATATGAGAGAGTGAAGTGTGCGGCTGCGCGG 300  
Db 56997 TCGTGGGCGCATTCGCGAGCTGACGATATGAGAGAGTGAAGTGTGCGGCTGCGCGG 57056  
Qy 301 GAGGATCAACGCTCGGCTCGCAAGCTCGTCAACCCGTCGTTACGTCAACGCGCATCAGC 360

Db 57057 GAGGATCAACGCTCGGCTCGCAAGCTCGTCAACCCATGTTTACGTCAACGCGCATCAGC 57116  
Qy 361 CTGCGGCGCGCAATATCAGCGACCGTGCACACAGCTGCTGAGTCTCGCTCCGAGCA 420  
Db 57117 CTGCGGCGCGCAATATCAGCGACCGTGCACACAGCTGCTGAGTCTCGCTCCGAGCA 57176  
Qy 421 GAGAGTTCGACGTTGTGTCGGGATTAACGAGAGGAAATCCGATGCGCGCATCAGCCT 480  
Db 57177 GAGAGTTCGACGTTGTGTCGGGATTAACGAGAGGAAATCCGATGCGCGCATCAGCCT 57236  
Qy 481 CTGTTGAAGTTCGCGCGAGTGTGACGAGAAATTCGCTCGCTTGGCTCGGCGACTGG 540  
Db 57237 CTGTTGAAGTTCGCGCGAGTGTGACGAGAAATTCGCTCGCTTGGCTCGGCGACTGG 57296  
Qy 541 CCGCGCTCGCGCGTGGTGTGTCGCCAGGTCGATGAGAGAACCAACCTGCTCCG 600  
Db 57297 CCGCGCTCGCGCGTGGTGTGTCGCCAGGTCGATGAGAGAACCAACCTGCTCCG 57356  
Qy 601 TCCGTCAACGAGGAGCTGCGCTGCTCCATGACGTCCTGATGAGCGCGAGAGAACCG 660  
Db 57357 TCCGTCAACGAGGAGCTGCGCTGCTCCATGACGTCCTGATGAGCGCGAGAGAACCG 57416  
Qy 661 CTCGAAATGACGTTTGACGATGCTGCTTCAAGCGCGAGCCGACGACGAGCTGAGC 720  
Db 57417 CTCGAAATGACGTTTGACGATGCTGCTTCAAGCGCGAGCCGACGACGAGCTGAGC 57476  
Qy 721 ACGAAGAGCTGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
Db 57477 ACGAAGAGCTGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57536  
Qy 781 TACCTTATCGGCTTGCCTGCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
Db 57537 TACCTTATCGGCTTGCCTGCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57596  
Qy 841 AAGCGCGAGCCCGGCTCATAGAGAACGCGCTCGATGAGTGTCTCCGCTTGCACAAATATC 900  
Db 57597 AAGCGCGAGCCCGGCTCATAGAGAACGCGCTCGATGAGTGTCTCCGCTTGCACAAATATC 57656  
Qy 901 CTCGAAATGAGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
Db 57657 CTCGAAATGAGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57716  
Qy 961 AAGAAAGGAGAGTGTCTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 1020  
Db 57717 AAGAAAGGAGAGTGTCTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAGCTGTATTC 57776  
Qy 1021 TCCAGGCCAGACGTGTGTGATGTGACCGGAGACAGGCGCGAGCCTGCGGTAGAC 1080  
Db 57777 TCCAGGCCAGACGTGTGTGATGTGACCGGAGACAGGCGCGAGCCTGCGGTAGAC 57836  
Qy 1081 GGGCCCAATGTCTGCCCCGGGGGTGTCCTGCTGCTGCTGAGGCGGAGATGCGCGTGGGC 1140  
Db 57837 GGGCCCAATGTCTGCCCCGGGGGTGTCCTGCTGCTGCTGAGGCGGAGATGCGCGTGGGC 57896  
Qy 1141 ACCATCTTCCGTAGGTCCCGAGATGAGCTGAAAGAACTCCGCTGTTTGATATCAC 1200  
Db 57897 ACCATCTTCCGTAGGTCCCGAGATGAGCTGAAAGAACTCCGCTGTTTGATATCAC 57956  
Qy 1201 CCGCGTTCCGGAACATGATCACTCAACGTCACTTTGAGACCTCCAAAGCTGATTA 1259  
Db 57957 CCGCGTTCCGGAACATGATCACTCAACGTCACTTTGAGACCTCCAAAGCTGATTA 58015

RESULT 11  
AP005958/c  
LOCUS AP005958 302650 bp DNA linear BCT 28-MAR-2003  
DEFINITION Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 24/31.  
ACCESSION AP005958 BA000040  
VERSION AP005958.1 GI:27354550  
KEYWORDS  
SOURCE Bradyrhizobium japonicum USDA 110

ORGANISM	Bradyrhizobium japonicum USDA 110
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.
AUTHORS	1 Kaneke, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110
JOURNAL	DNA Res. 9 (6), 189-197 (2002)
MEDLINE	22484998
PUBMED	12597275
REFERENCE	2 Kaneke, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL	DNA Res. 9 (6), 225-256 (2002)
MEDLINE	22485002
PUBMED	12597279
REFERENCE	3 (bases 1 to 302650)
AUTHORS	Kaneke, T.
TITLE	Direct Submission
JOURNAL	Submitted (20-NOV-2002) Takakazu Kaneke, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
MEDLINE	(E-mail: kaneke@kazusa.or.jp.
PUBMED	URL: http://www.kazusa.or.jp/rhizobase/
REFERENCE	Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)
AUTHORS	Location/Qualifiers
TITLE	1. .302650
JOURNAL	/organism="Bradyrhizobium japonicum USDA 110"
MEDLINE	/mol_type="genomic DNA"
PUBMED	/strain="USDA110"
REFERENCE	/db_xref="taxon:224911"
AUTHORS	31. .1110
TITLE	/gene="blr6272"
JOURNAL	31. .1110
MEDLINE	/gene="blr6272"
PUBMED	/note="Arac family
REFERENCE	ORF ID:blr6272
AUTHORS	/codon_start=1
TITLE	/transl_table=11
JOURNAL	/product="transcriptional regulatory protein"
MEDLINE	/protein_id="BACS1537.1"
PUBMED	/db_xref="GI:27354551"
REFERENCE	/translations="MENLBSLVESFGLGGLPVASGRDAKGRAMFTDLOAAMPID LQAEPRDHLQAOAI PERQWRIRIDRPEAIEBITTSRMEDFSSSDATTED RYFVGIALKATRAKLTDRQVI FDGTMPTGLVSAKSPRLAQAQAPCAFPHISA DHFPASPAEGIGDLVLDRLDPLAELARLTHEGAADREFRCVQOTLAFHILAE PPRANALPKMLRREVOYIADNPFRCISLSLAVAGISRHHPAAGORTATGYPR EYLINRIRIHAKTILATTGRPLAETILAVGFSQAHFSTVFKRISGSAARRRLASKA BPLAEALPRRRPADRDMMASAAA"
gene	complement (1428. .3794)
CDS	/gene="blr6273"
	complement (1428. .3794)
	/gene="blr6273"
	/note="ORF ID:blr6273
	probable biotin sulfoxide reductase"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS1538.1"
	/db_xref="GI:27354552"
	/translations="MDRYATATWYRNKQMMDDTIGPPDGLDLSGFKRPTSHWGVF ARNGAGLEVRAYACDPDPNGITDNFGALRHQARIQAQPAIRRGWLERGPGDPRGR DERFVSWEADLDLDELGRIRDTGPAPVFGSGWSAGFPHASQVHRFLTA MGVYRVSNSVSGASVLLPQILAGYEDITKKNVMEQIATETDVLAFGQALNKS MVGASISKVERSAMEARKGCEFIIVSPLAEDLPVAGAEWMTCVSTDTALMLG IVTTLSEKLDQAFLDRTYEGWPFPLRITGSDGPPSAEWAALISGVDATTTL ARLAKRALITSHSLQRAHEGQPVMMGMVLAALGQIGLPGGYAIVSLAIGYIG
gene	complement (7993. .8880)
CDS	/gene="blr6274"
	complement (3898. .4947)
	/gene="blr6274"
	/note="ORF ID:blr6274
	putative dehydrogenase"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS1539.1"
	/db_xref="GI:27354553"
	/translations="MTGASDTVAIGADLRPPADLLRAGGFTADQADEAAMLVAN LRGASHGVLRIPRYVEMVEGLINPAEPRQVSGKAVAVIADQARAGAVMSLAAR KASBLADTHGICWCAVRSI SHAAGIYTAHOVAQAGRTGYMTASKPLMTYHGSAG VSTNPILAIAPLTPNDRPLLDMSGTASVALGKIMARDAQPIPVGAIVDNGQATTD PREAKAVLPMAGPPKSGSLMI EVLASVLTANPLIAYALVHGDPGSGLVAVLDPSA FGDNFTLEVDRLGAIKSLPPAGADSVYLPGBRGFEEMERRSRAGITPLRAGTVSRL ALASKIANVAPEGVA"
gene	complement (4950. .5873)
CDS	/gene="blr6275"
	complement (4950. .5873)
	/gene="blr6275"
	/note="ORF ID:blr6275
	hypothetical protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS1540.1"
	/db_xref="GI:27354554"
	/translations="MSAPLSQPTREIPMOEKHTRSDGLKLSAVLHVPDKRPGQRL PAFVILHGFSGKDSHAEIOARMLEQGYALAFDFCCSGESGEARQVCFQVAD AKNALTPLAERBEVDPRIGVGHSGAIVYAGVDRPACVSGSGMDGSEKFR GOHPREMAKFTSLLEKGRKOKETGOSMMISRDAVIPBHLRKNLSPLAIMEIPV ETAMSMYPRADENVYANPRPILFHLNNTITPTBOSIMFEAGQAPALVLTGT SHPLAEDAPRTKAIKQWLDKFPSPPLGA"
gene	complement (5910. .6812)
CDS	/gene="blr6276"
	complement (5910. .6812)
	/gene="blr6276"
	/note="ORF ID:blr6276
	unknown protein"
	/codon_start=1
	/transl_table=11
	/protein_id="BACS1541.1"
	/db_xref="GI:27354555"
	/translations="NPTVAVGHYKINTTALVAVRPATETLGIQITYTDEGQWLS SNGRAELVILRSHENATHTTGLEALTVDAVREASRVESAGCRVTSKSPSLDCAAG VTFDPEGLRFEVHTPTDQIYGRYPTPHGVPNMDHLNLTSPDPIATROQLERI LIRSIRATDSSLMVYGNRQHHIYGVVGRKGLHLYFEFEFNDYKLDGLDRFD ROMLNGPRHRPGDNTVAYVYDSSGAMIECGSMALLDDQGFVNVNITNLERPGNVR CMNVGTGPALEMRQPHRPTAIT"
gene	complement (7080. .7853)
CDS	/gene="blr6277"
	complement (7080. .7853)
	/gene="blr6277"
	/note="GntR family
	ORF ID:blr6277"
	/codon_start=1
	/transl_table=11
	/product="transcriptional regulatory protein"
	/protein_id="BACS1542.1"
	/db_xref="GI:27354556"
	/translations="MILGSRGMAARRMRBRVYAGSTRADVLAHRADIISCTLK PCAKRFEALRITAYVSGTTRALSRLVABGLVYABORGFVYAVSIDLNDLITV RVLIREVCVLAIKNGDAWEADITIGAFHRMDRLQNRGNTYLSSENGKLHDGHS LVAAQSNLLALEIRKQLEFARHRYRMSQFRTKRAVDVEHKMTDSVVAADAKAE ELIEHIRETNTYIKHGHLPVTSDEBRPRTSRVAE"
gene	complement (7993. .8880)

```

CDS
/gene="11ve"
/complement (7993..8880)
/gene="11ve"
/note="ORF_ID: b116278"
/codon_start=1
/transl_table=11
/product="branched-chain-amino-acid transaminase"
/protein_id="BAC51543.1"
/db_xref="GI:27354557"
/translation="MVPTNSISHSRTWTFPRGQWHEGVPIMGRPTGLMALTIPD
GAPAFEGVMPDLDRHCAVNSAVNFGKPVVDLTLGLVREGIAPVAAETIRP
MYWPONGIGGAVLEDPETTDWCLCYSEPIEPVGSATLSPFPAKAPVAKAS
CLYPSRSALREASRFQNCIMDLNTEFNGSNATPMKDDVFPAPANGFLNG
ITRORVILELRGDIIVLEATIRYSDFVSADEIFSTGNFAPVIRIDRQLRGRF
YARAKLYWDFAHGKLA"
complement (8909..9784)
/gene="b116279"
/complement (8909..9784)
/note="ORF_ID: b116279"
/hypothetical_protein"
/codon_start=1
/transl_table=11
/protein_id="BAC51544.1"
/db_xref="GI:27354558"
/translation="MRVMTMRCREIGIADLADLTTRGPTSRDYIDGLRRLS
VRETPDSFRRGYMLEHDPARVAGALOLITRANGGSISTRCNLSMWVEBAPSYAP
MLTYAQRHKNVTYVNIISPAETWYVIGGPRRRCGLSPSPALSPAPAMVEV
RQDQADIDLEADVILARHAAAGLCLICRASGRAPFVLDPMRVRIURLAMOM
IYCPDVAVVYVAGAILGRFLIRGRISIALDANGPVRGLVGIYRBIQRKYFKPHCP
RLADLSDSELVIGP"
complement (9781..10938)
/gene="b116280"
/complement (9781..10938)
/note="ORF_ID: b116280"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAC51545.1"
/db_xref="GI:27354559"
/translation="MTVLIDVSDSEHLLSRAPNGARAMASPGNRACLDIGLNNM
SDAALMSTERQDLFLDLAAAGRLCVRLHFYMEATPREMERDVRVRYRGIIDLNR
RLDGVITVGAEPRAASLTDEPWTTFPAIADMARNTVSSECLAVGAVLHDGVT
RHKLPTGCFVFAOTKRHHPLMODVPTPRIIPARMVVEBDSGVSVLMSAE
AGVPCFYQKQKNSLFTVHOGHPREYTGSLSEYRDMKRFAGEHEVPTTIPRGLNK
GAETLLAFRQKALSDROPFLPADPAPDLTRDLNVHLPAOHLYRNWLOVMSQPA

Query Match      12.5%  Score 157.4; DB 1; Length 302650;
Best Local Similarity 49.4%; Pred. No. 6,4e-17;
Matches 556; Conservative 0; Mismatches 551; Indels 18; Gaps 5;

QY 49 TTCAAGCCGTTGCGCGCTGGGATCGCGAGAGACCCGTTCCCGCGATGAGCGCTTGA 108
Db 294417 TTCAATCCGCTGTCGCGAGACTTCATCCGCAACCCCTATCCGATTAAGACGCGCTGGC 294358
QY 109 GAGCGAACCCCATTTTACTGCGGATGAGAGCGCGCTCTGCTGCTCAACCCGATACAC 168
Db 294357 GCGATGATCCGATCATGTGACGCGCTTGGCCAGTTT--GTGCCAGTGGCCACGCC 294301
QY 169 GACGTGTGCGGCGTTCGCGAGACGCTTCGCGTCACTGCGAGAGAGTGGATTCG 228
Db 294300 GATGTCAACCTCGTATCGCGACAAAGCGCTTCGCGCAAGATTTGTCGAGCGGCTCCAG 294411
QY 229 AGCGCGAGTACTGCTCGGCATTCGCGAGC--TCAGCGATATGAGAAGTACGATTCG 285
Db 294240 CGCGCGTACAGCAAAAGATATGACGAGCGGCTGTTCCGAGCATGACCACTGATG 294181
QY 286 TTCGGCGTCCGCGAGAGATACGCTGCGGTCCGCAAGCTCTGCAACCCGCTGTTACG 345
Db 294180 CTGACGCGCGATCCGCGACCAACCGCGCTGCGGCGCTTGTGTGTAAGGCCCTTACCC 294121
QY 346 TCACGCGCATGACTGCTGCGCGCGCAAAATACAGCCACCGTGCACGACTGCTCAT 405

```

```

Db 294120 GCGCGCGCGCTCGAGCATCGCGCGCGCGATCCAGAGATGCTGATGAGGCCATCAT 294061
QY 406 GCTGCGCTCCGAGCAAGAGAGATTGAGAGTTTGCGGGATTACCGCGAGGAATCCCGATG 465
Db 294060 GCCGTGATCGACCGCGCGCATGAGACTGTATCGAGATTTTCGCTCCGCTGCGCTC 294001
QY 466 CGCGCGATCAGCGCTGTTTGAAGATTCCGCGCGAGTGTGACGAGAAGTTCGTCGCTTC 525
Db 294000 ACCATCATCTCGCATACCTGTTGGAAATCCCGAGATCATCCGAGTCTT-----CTAC 293947
QY 526 GCTTCGCGACTGCGCGCGCGCTCGCGCGTGGATTGGTTCGCCAGTGCATGAGAGAC 585
Db 293946 AAGAGCTGCGCGACGCGCGCGCGCTGCTGCACCCCGGCGCTTCACCGCGAGAGATC 293887
QY 586 AAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645
Db 293886 GCGAAGGCGCAACGCGCGCGCATGATGATGATGATGATGATGATGATGATGATGATG 293827
QY 646 CGCGCGAGGAACCGCTCGAATAATGACGCTTGAAGATGCTGCTTCAAGCGCGAGCGAG 705
Db 293826 CGCGCGCGCAATTC--CGCGAGACCTTACCAACGCTGCTGCGAGCGCGAGAGAC 293770
QY 706 GCGACAGGCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
Db 293769 GCGAACAAGCTACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 293710
QY 766 ACCGATACAGATCTACTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825
Db 293709 CACGAGACCAACGCTGATCTGATGCGCAACGCGCTCTTGGCGCTGATCGCAACCGGAC 293650
QY 826 GCGCTCGAGCTGTGGAAGCGCGAGCGCGGCTCATGAGGAACGCGCTGATGAGTGTCTC 885
Db 293649 CAGCTCGCGCTGCTCAAGAGCGCGCGCGAGCTGATGCTGATGCTGATGAGAGTCTTG 293590
QY 886 CGCTTCGACATATCTCAGATAGAACTGTGCGCTTTCGCGAGAGAGAGAGAGAGAGAG 945
Db 293589 CGTACAGATCTCGGTGAGATGACGCGCGCGCTGACGCTGAGAGATGACAGATCTC 293530
QY 946 TCGCGGCGATGATCAAGAAAGGAGAGATGCTTCTTCTGATCCGAGCGCTGAGAG 1005
Db 293529 GCGCGCAGAAATATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 293470
QY 1006 GATGGAGTGTATCTCTC--AGCGCAGAGCTGTTTGAAGTGTGCGAGCGGAGACGCGCG 1062
Db 293469 GATCGCGGCTGTATCCCGATCGCGCGAGCGGCTGACAGTCAACCGCTCCCAAGCTCAAG 293410
QY 1063 AGCTTCGCTACGAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db 293409 CCGCTGCTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 293350
QY 1123 GCGGAGATCGCGGTGGGACCATCTTCCGTAGGTTTCCCGGAGATG 1167
Db 293349 GCCGAGATCGCATGCGCACGCTGTGGCGCGCGCTGCGCGAGATTCG 293305

RESULT 12
SERCP450A
LOCUS      SERCP450A      3078 bp      DNA      linear      BCT 26-APR-1993
DEFINITION Saccharopolyspora erythraea ORF 1 gene, partial cds; cytochrome
P-450 gene, complete cds; ORF 2 gene, partial cds.
ACCESSION M81110
VERSION    M81110.1 GI:152682
KEYWORDS   cytochrome P450.
SOURCE     Saccharopolyspora erythraea
ORGANISM   Saccharopolyspora erythraea
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.
REFERENCE  1 (bases 1 to 3078)
            Andersen, J.F. and Hutchinson, C.R.
            Characterization of Saccharopolyspora erythraea cytochrome P-450
            genes and enzymes, including 6-deoxyerythronolide B hydroxylase
            J. Bacteriol. 174 (3), 725-735 (1992)
JOURNAL

```



MEDLINE 92121109  
PubMed 1732208  
Original source text: Saccharopolyspora erythraea (library: NRRL  
COMMENT 2338) DNA.  
FEATURES  
source  
1.3078  
Location/Qualifier  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1836"  
/tissue\_11b="NRRL 2338"  
<1.859  
/note="ORF homologous to Escherichia coli and Salmonella  
typhimurium dnaB gene; putative"  
/codon\_start=2  
/transl\_table=11  
/protein\_id="AAA26482.1"  
/db\_xref="GI:152683"  
/translation="AGGGRGHRAPAPAPAPVPRVRRRRRRRRRVVAVAGAVHAG  
RYAQAADLRADGAPLRGPGSGRRAARAVGHRHRRARRRRRGGGPGDR  
DLGDPAPHQRRPVGCDRPGRRRORRGATLPEVLRRLPARGHDRREGHOR  
ARGHHOHLRRGAGHLDGRDPGNOPRVROGAGGCGPVAGRAQHNLGALGHG  
PGARBAAGARHPRAVQRLRLVDRERAGRAEAGARRGVLRDGAVSRRGAASPLRA  
PPRLRGS"  
949..2166  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450"  
/protein\_id="AAA26483.1"  
/db\_xref="GI:152684"  
/translation="MTTGEVPLDLAFDAFADRRNRVYRMKEEPVQIRITVGLDAM  
LITRYEDVKQALLDPRIAKDFGRTQOIIIEKRLADARRGFSFDGPHMLNTPDHT  
RLRLVAVFCTARVGLAPRIEOTDLDLRLADARRSEVDLDERAPPLPVIIEIM  
GVEDSRDDFRSTWTVLDGSPERAOQASVAMERYITELIKKTEPDDLLTALLE  
AVEDDLRSEGLIANVFLVAGHETVNLIGCVLSLGNPDOLALRNDPSLLPG  
ALEBTLRSES PVANGTFRRTAEVAFVGVAVGGLSLTAIRAFLETPGPHAST  
ITRETTGHAVFHGHIHFCVGAALARLEAQIAGRLIERPPDLRMASPDLLMRSEVL  
NRGLKLPVRPQA"  
complement(2365..>3078)  
/note="ORF 2"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA26484.1"  
/db\_xref="GI:152685"  
/translation="LIJAVATGSLFVLSOYLIVRGHSAVVAGMAALPLAASVGS  
LGRAPRIRGACTVTRAVYTAGCVTAALGPPSGPHIIFILLCFPGTGPAP  
ATSLAGVADPADRAGSLANDTHOQLGAFVAVGGLSLTAIRAFLETPGPHAST  
SLAATLSFADERTSALADAPRLAFOAQSATMTAGLACALAGAAVVMILSRGRASA  
VALTVSATARRNLRAAGT"  
BASE COUNT 463 a 1060 c 1137 g 418 t  
ORIGIN  
Query Match 12.2%; Score 154; DB 1; Length 3078;  
Best Local Similarity 49.1%; Pred. No. 5.3e-16;  
Matches 569; Conservative 0; Mismatches 540; Indels 51; Gaps 4;  
QY 45 CGACTTCAGCGCTTGGCGCTGGGTAAGCGGAGCGGTTCCCGCGATGAGCGCT 104  
DB 965 CGACTTACTGCTTTCAGCAGCGCTTCCGCGAGGACCGGCAACCGGTACGCGGAT 1028  
QY 105 GAGAGGCGAC-----CCCATCTTCTACTGAGTGAAGCGCGCTCTCGGTCCTCAC 158  
DB 1029 GGGCGAGGAGCGGCTGACGCGATCGCAGCGTGAACGGCGCTGACGCGGTATATC 1086  
QY 159 CCGATTCACGACGCTGTGCGCGGTTCGCGCAGCAACCGCTTCCGCGTATGAGAGA 218  
DB 1089 CCGATTCAGGAGCGTTCACAGCGCGCTGCGACCCACAGATCGCCAGACTTCGCGCG 1148  
QY 219 GTGGGATGAGAGCGGAGTACTCGTGGCGCAATCCCGAGCTCAGCGATATGAGAGA 278  
DB 1149 CACCCGACGATTCATGAGAGAGCGTGGCCGACGCGGAGCGCGGGGTTCTCGCC 1208  
QY 279 CGGATTGTTGGGCTCGC-----CCGAGGATTCACGCTCGGGTCCGCAAGCT 326

DB 1209 CGACCTGGGCCCGCATGCTCAACACGACCGCCGAGACCAACCGGTTGGCAAGCT 1268  
QY 327 GGTCAACCGCTGCTTACGTACAGCGCCATGACCTGTGGCGCGGAAATACAGCGAC 386  
DB 1269 GGTGTCAAGGCGTTCACCGCGCGCGGCGGAGGCGCTGGCGCGCGATGACAGAT 1328  
QY 387 GGTGACAGCGTCTGAGTGTGCTGCTCGGACCAAGAGAGTTCGAGTGTGGGAGATTA 446  
DB 1329 CACCCACGACCTGTGTGACCGCGCTGGCGGCGCGTCCAGATGATTCGATGACAGATT 1388  
QY 447 CGCGAGGAGAAATCCGATGCGCGCGCATGACGCTGTGTGAAGTTCGCGCGAGATGTA 506  
DB 1389 CGCGTTCGCGCTGCCCATCAAGTATCTCGAGCTGATGAGCGCGTGGAGATCGCGGG 1448  
QY 507 CGAGAGTTCGCTCGCTTCCGCTCGGCGACATGCGCGCGCGCTCGCGCTGGGTTGTC 566  
DB 1449 CGACACCTTCGCG-----TCGTGACCAACGCTCTGTGTCAGCGGCTCGACCGGA 1499  
QY 567 CCAGTTCATGAGAGACCAAGACCTGTGCGCGCTCGACCGAGGGGCTCGCGCTGCT 626  
DB 1500 GCGCAGGCGCCAGCGCTGCGTGGCGATGATGAGTACCTGACCGAGCTGATGCGCAAGAA 1559  
QY 627 CCATGACGTCTCTGATGAGCGCGCGAGAACCCGCTCGAAAAATGACGCTTTCAGATGCT 686  
DB 1560 GCGCACCGAGCCCGGTGAC-----GACCTGTGACCGCGCT 1595  
QY 687 GCTTACAGCCGACGCGCCAGCGACGAGCTGATGACAGAGAGCTGTGCGCTGTGG 746  
DB 1596 GCTGAGGCGCGTCTGAGACGCGGACCGGCTCAGCAGAGGCGAGCTGATGCGCATGTGTT 1655  
QY 747 TCGCATTCGCTGCTGCGACACGATACCAAGATCTTACCTTATCGGCTTGGCTGCTCA 806  
DB 1656 CCGTCTGCTGCTGCGCGCGGACAGACAGACGCTCACTGATCGCAACTGTGCTGTGC 1715  
QY 807 CCGTCTGCGCTGCGCGAGCGCTTGAAGCGCGAGCGCGGCTTCATGAGAA 866  
DB 1716 GCTGCTGGGCAACCCGACAGCTGCGCGCTGGCAACGACCGCTGCTGCTGCCG 1775  
QY 867 CGCGCTCGATGAGTGTCTCGCTTTCGACATATCTCAAGATAGAACTGTGCTTGGC 926  
DB 1776 CGCGATTCAGAGAGACCTGCGCTTACGAGAGCGCGGTGCGAAGCGGATTCGCGACAC 1835  
QY 927 CAGGAGAGCTCGAGTACTGCGGGGCGATCGATCAAGAAAGGAGATGTCTTCTCT 986  
DB 1836 CGCGAGCGGCTGCGCTTTCGAGAGCTGTATCCCCAGAGAGCTGTGTGCTGCG 1895  
QY 987 GATCCGAGCGCGCTGAGAGATGGAATGTAATCTTCAGGCCAGCGTGTGTGATGTCG 1046  
DB 1896 GCTCGGGCGGGGAACCGGAGCGGTGAGCGCTTCAGAGACCCCGCTTCGACATCAC 1955  
QY 1047 ACGGACACGGGCGGAGCTTGCCTACGTAAGAGCCCCCATGTCTCCCGGGGTGTC 1106  
DB 1956 CCGGAGAGACCAACCGGAGCGCTTTCGCGCACCGGATTCATCTTCTGTGGGGCGCG 2015  
QY 1107 CTTGCTGCGCTCGAGCGGAGATGCGCGTGGGACCAATCTTCCGTAGATTCGCCAGAT 1166  
DB 2016 CTTGCGCGGCTGAGAGCGCGCATGCGCTGCGGACGCTCTCGAGCGCTTCCGCACT 2075  
QY 1167 GAACTGAAGAACTCCCG 1186  
DB 2076 GCGGATGCGCGGTGCGCGG 2095  
RESULT 13  
AX697983 1233 bp DNA linear PAT 02-APR-2003  
LOCUS AX697983  
DEFINITION Sequence 7 from Patent WO03010193.  
ACCESSION AX697983  
VERSION AX697983.1 GI:29499037  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE 1  
AUTHORS Farnet,C.M., Staffa,A. and Yang,X.  
TITLE Genes and proteins for the biosynthesis of rosamycin  
JOURNAL Patent: WO 03010193-A 7 06-FEB-2003;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
source  
1. .1233  
/organism="Micromonospora carbonacea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:47853"  
BASE COUNT 225 a 452 c 377 g 179 t  
ORIGIN  
Query Match 11.9%; Score 149.4; DB 6; Length 1233;  
Best Local Similarity 49.7%; Pred. No. 3.6e-15;  
Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;  
Qy 288 CGGGCTGCGCGCGAGATCAAGCTCGGATCGCAAGCTCGTCAACCCGCTGTTACGTC 347  
Db 261 CGGCATCGACCCGCGCGGACGCGCGCTGCGCAAGCTGAGACCAAGGCGTTACCC 320  
Qy 348 ACGGCCATCGACCTGCTGCGCGCGCAAAATACAGCGCACCGTCAACAGCTGCTGATGC 407  
Db 321 CGCGCGATCGCGCGAGATGAGCTCGGATCGGCGAGATCACCGCGTCTGACCA 380  
Qy 408 TCGCTCGGAGCAAGAGAGTTGACGTTGCGGGATTACGCGGAGGAATCCGATCG 467  
Db 381 GGTACGCAACGAGCGGATGACATCGACGCACTCGGATACCGCGTCCGATAC 440  
Qy 468 CGCGATCGCGCTGTTGAAAGTTCCGCGCGAGTGTACGAGAAATTCGTCGCTCG 527  
Db 441 GGTATCGCGAGCTGCTCGGATTCGCAACCAAGATCACGAGATTCGCGAGTGGT 500  
Qy 528 CTCGCGACTGCGCGCGCTGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 587  
Db 501 GACATCATCTCTGACCAAGAGAGGCTGAGATCCCACTCCGAGCACTTACCGA 560  
Qy 588 GACCTGTGCGGCTCGTCAACGAGGCGCTCGGCTCCATGACGTCCTGATGAGCG 647  
Db 561 GACGCTGCGCGCGCGATGAGAGTGGTCCGAATTCCTGACCGCGAGTCCCA 620  
Qy 648 GCGCAGAACCGCTCGAAATGACGCTTGAAGTCTTCAAGCGCGAGCGCGA 707  
Db 621 GCGCGCGAACCG---AAGAGACGCTGATGACGCGCTGCGGCGAGGATGAGCG 677  
Qy 708 CAGCAGCTGAGCAGAGAGAGCTGCGCTGCTGGTGGATTAATGCTGCTGAGC 767  
Db 678 GCGCAAGCTGACCGAGAGAGTGTCAACATCGTGGCTGCTGCTACCGCGGCA 737  
Qy 768 CGATACAGATCTTATCGGCTGCTGCTCAACCTGCGCGCTGCGCGCGAGCG 827  
Db 738 CATCTCAGCGCGCAGCTGCTGAGCACTGTTCTGCTGCTGAGAGCAACCGCA 797  
Qy 828 GCTCAGCTGTGAGAGCGCGAGCGCGCTCATGAGAAAGCGCTCGATGAGAGTCTCG 887  
Db 798 AAGCGCGCGCTGCGCGCGAGCGAGCTGCTGCGCGCGCTGCTGAGAGAGCTGCG 857  
Qy 888 CTTCGACATATCTCTGAGATGAGAACTGCTGCTTCCGAGGAGAGCTGAGTACTG 947  
Db 858 CT---ACGGTCTCCGTTCACTGATCTTCGAGATCTTGAACGAGGACACGATCT 914  
Qy 948 CGGGCATCGATCAAGAAAGGAGATGCTTTCTCGATCCGAGCGCGCTGAGGA 1007  
Db 915 CGGCAACCCATGCGCAAGGCGCAGATGATGCTCCGATGCTGCTCCGCAACCGGA 974  
Qy 1008 TGGACTATTTCTCAGGCGCAGAGCTGTTGATGAGCAGGAGCAGGCGCGAGCT 1067  
Db 975 CACGAGGATTTCAAGAGACCGGACACCTTGACATCGAGCGGATGAGAACACACT 1034  
Qy 1068 CGCGTACGTTAGGCGCGCGATGCTGCGCGCGGCTGCTGCTGCTGAGGCGGA 1127  
Db 1035 GCGCTTGGCGCAGCGATCACTGCTGCGCGCTGCTGCTGCGCAGGCTGAGGCGGA 1094

Qy 1128 GATCGCGTGGGACCACTTCCGATGTTCCCGAGAT 1166  
Db 1095 GGTCTCTCAACGAGACGCTCGACAGTTACCGAGTT 1133  
RESULT 14  
AX697977/c 60196 bp DNA linear PAT 02-APR-2003  
LOCUS AX697977  
DEFINITION Sequence 1 from Patent W003010193.  
ACCESSION AX697977  
VERSION AX697977.1 GI:29499034  
KEYWORDS  
SOURCE  
ORGANISM  
Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporinae; Micromonosporaceae; Micromonospora.  
REFERENCE 1  
AUTHORS Farnet,C.M., Staffa,A. and Yang,X.  
TITLE Genes and proteins for the biosynthesis of rosamycin  
JOURNAL Patent: WO 03010193-A 1 06-FEB-2003;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
source  
1. .60196  
/organism="Micromonospora carbonacea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:47853"  
BASE COUNT 6934 a 20042 c 23822 g 9398 t  
ORIGIN  
Query Match 11.9%; Score 149.4; DB 6; Length 60196;  
Best Local Similarity 49.7%; Pred. No. 1.8e-15;  
Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;  
Qy 288 CGGGCTGCGCGCGAGATCAAGCTCGGATCGCAAGCTCGTCAACCCGCTGTTACGTC 347  
Db 3601 CGGATCGACCCGCGCGGACGCGCGCTCGCAAGCTGAGACCGAGGCTTACCC 3542  
Qy 348 ACGGCCATCGACCTGCTGCGCGCGCAAAATACAGCGACCGTCAACAGCTGCTGATGC 407  
Db 3541 CGCGGATCGCGCGCGAGATGAGCTGCGGATCGGCAATCACCGCGAGCTGACCA 3482  
Qy 408 TCGCTCGGAGCAAGAGAGTTGACGTTGCGGAGTTACGCGAGGAGATCCGATGCG 467  
Db 3481 GGTACGAGCAGAGCGCGAGTGAATGATCGACGAGGACCTCGATACCGCGCGTAC 3422  
Qy 468 GCGCATAGCGCTGTTGAAAGTTCCGCGGAGTGTGACGAGAAATTCCTGCTTGG 527  
Db 3421 GGTATCGCGAGCTGCTGCGATTCGCAACCAAGATCACAGAAATTCGCGAGTGGT 3362  
Qy 528 CTCGCGACTGCGCGCGCTGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 587  
Db 3361 GACATCATCTCTGACCAAGAGGCTGAGATTCGCACTCCGCGAGCTTACCGA 3302  
Qy 588 GACCTGTGCGCTCGTCAACGAGGCGCTGCGCTGCTCATGACGCTCGATGAGCG 647  
Db 3301 GACGCTGCGCGCGCGCGATGAGAGTGGTCCGAATTCCTGATACCGCGAGATGCCCA 3242  
Qy 648 GCGCAGAACCGCTCGAAATGACGTTGACGATGCTGTTCAAGCGCGAGCGCGAGCG 707  
Db 3241 GCGCGCGAACCG---AAGAGACGCTGATGAGCGCGCTGCTGCGCGAGGATGAGCG 3185  
Qy 708 CAGCAGCTGAGCAGAGAGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 767  
Db 3184 GCGCAAGTGAACGAGAGAGTGTCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3125  
Qy 768 CGATACAGATCTTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 827  
Db 3124 CATCTCAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3065  
Qy 828 GCTGAGCTGTGAGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 887  
Db 3064 AAGGCGCGGCTGCGCGCGCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3005

QY	888	CTTCGACAAATATCCCTCAGAAATAGAAACTGTGGTTCCGACGAGCGACGACTGGAGTCTG	947
QY <td>948</td> <td>CGGGGCATCGATCAAGAAAAGGGAGATGCTCTTTCTCTGATCCGAGCGCTTGAGAGA</td> <td>1007</td>	948	CGGGGCATCGATCAAGAAAAGGGAGATGCTCTTTCTCTGATCCGAGCGCTTGAGAGA	1007
Db	2947	CGGGCACCCCATGCGCAAGGGCGCAATGTAATGCCCTCGATATCGCTTCGCGAACC	2888
QY	1008	TGGGACTGTAATCTCTCAGGCGCAGAGCTGTGGATGTGCGACGAGGACACGGGCGCGAGCCT	1067
Db	2887	CACGAGAGTGTTCACGAGACCGGAGCACTTGAGACATCCGACGCGAGTGAACAAAGCACTT	2828
QY	1068	CGCGTAACGTTAAGAGGCCCCCATGTTGTCCCGGAGGTGTCCTTGCTGCCTCGAGGCGGA	1127
Db	2827	GGCGTTTCGGCCACGCGCATTCACATCTGCTGGGCGCGTTTCTCTGCGCCAGCTGAGGCGAA	2768
QY	1128	GATGCGCGTGGGCAACATCTTCCTCGTAGGTCGCCCGAGAT	1166
Db	2767	GGTCTTCTTCAACAGACGCTCGACCACTTACCGAGTT	2729

**CDS**

CDS  
6369 . 7730  
/note="ORF7"  
/codon\_start=1  
/transl\_table=1  
/product="protein kinase"  
/protein\_id="CAD1.9078.1"  
/db\_xref="GI:19572311"  
/translation="MERLI35ALPDGSRVGPMPFIEGPAAGCTGYGAVYVRAHRYVRCR  
SVALKARYPNDLRFERRBAELIERVRANYPRLHGHTWAGQGEKHPITVYQWQVQ  
VRLTOMNKEKRIARIEVLRLVLAQVARKLEKTHMGKSHRVKSNVLAPEBGLSLMD  
FGCGTWEBAALITGGLIAPGTAKVAREALRFHNNHNNENADVAQATPDADVALGVT  
AYRLCTGYPPPATPEPSGVGDERETDEVKKPSPQVAPLPALAEOLILMLRNPNPKR  
GRABELAMEALARAAGAEADDEPMNPNHSVSEITGRVSVQREMGCPVALVLA  
VGLLIMGSGINTGRVAPAPASBMGTGADVADAVALPVSSARAEPSGNTLMD  
KEPLPGQRPPCPRRFQISIGCGCMLVPEQSSPGGSEYEMKKAQCSVPVAAPRPRTS  
DPQ"  
complement(7849 . 8577)

CDS

```

/codon_start=1
/transl_table=1
/product="methyl transferase"
/protein_id="CAD19079.1"
/db_xref="GI:19572312"
/translation="MSQNIYDNPSPFESYSOLARSLKGLAGDEWPAIRLRLDVIDL
RVYDLGCGFCGFCWARKNDAAHYGLDLSKMTLRKATSGDITAYQADLRVYL
PGCAFPLDSFLAIHYIELLEPLATYTHRLTGGWVPFIEHPIFMASSQPMWNG
EGHRTWPFDSITQLEGKRTTHMLAEVVKYIKHTLTINALLEQGFTHRIEWMPTDA
QLSHAPLEAERNDPMLVLVSAQR"
complement(8622..8879)

```

**CDS**

```

//nucleo_start=1
//codon_start=1
//transl_table=1
//product="acyl-CoA binding protein"
//protein_id="CAD19080.1"
/db_xref="GI:19572313"
/translation="MALEDPPRSAGQEVKALKTPSPNDRLLELYSLFKQATEGDVQOK
PPGLIDVPERKLYAMVGRKGLGRSMAQQVALLVERLNG"
8961..10070

```

**CDS**

```

/codon_start=1
/transl_table=1
/product="potassium channel beta chain"
/protein_id="CAD19081.1"
/db_xref="GI:19572314"
/translation="MVYLICRSRQLRMILINGSNRVRFAMGCHAPRRMHPHILGRS
GLAMSEISYGNMLTHGSOVEBAALVACVRALDGIITPTADVYAGTRAEVYGRAL
OGCRREGIEFTKYKYPGPGPNDRGISRTHLHISGLIRLKTETVDDLYOHRVDT
ETPLRETHQAISDVVRQGGKHHYIVSGSWTLDQLREBGLKSRQNGFQVLSVQPOYSNM
RVIREAEVPAQELVGQVQVYVSPAQSVLTGKRPFGKRPPEGSRAIDTEKGGASFIRK
LNDLVLTFTVQOLTVPVAQEGISLAQLAAVAVLQNPVSSAIIIGASRPEQYKENVKAG
VKLSSEARHRIIDLGPILVERNPCKTLSPRRKP"
10428 . . 10862

```

**CDS**

```

/codon_start=1
/transl_table=1
/product="ribosome binding factor"
/protein_id="CAD19088.1"
/db_xref="GI:19572315"
/translation="MTSEFKORRRAPRPGRESSLSLSEAOALPAHHLRVOSTFOEVA
RLPFGELSDLTLEGVSTSPFLTPDGNNAIGFLPPEVAGSGAPIALVALNAGSRI
PAQALGLNTRRVPHLRFVCGVAPPFQSFIPEDOSAEGES"
10859..12016

```

CDS

```

publication
Chem., 10.1074/jbc.M11738200
s 1 to 66808)

Submission
d (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
weieg, GERMANY

Location/Qualifiers
1. 66808

/organism="Stigmatella aurantiaca"
/mol_type="genomic DNA"
/strain="Sg a15"
/db_xref="taxon:41"
complement(4978..6631)
/note="ORF8"
/codon_start=1
/transl_table=1
/product="cytochrome P450 monooxygenase"
/protein_id="CAD19077.1"
/db_xref="GI:19572310"
/translation="MARSVSASRPPSPANPVAAFDLARLDLDAFYVDPPEPLYRKMRERER
PVHMRVPSLITLPMALDLCITMDRTPFSSDKRAEKGAYGAPLEFHHHTLIVNPENQ
PLHMRVRLVIGALTALDCTYMRPGRTLYVDRLDGLAKGADILIEPPAAALIPLEVD
GNLIDIPERNGPIRGKSLIILGALREPLRAEDKAGNAVTTEFDYIRLILVAQRRA
```

```

/protein_id="CAD1.9083.1"
/db_xref="GI:19572316"
/translation="MRPRLEGARPAVPIIAMAAPVAGAVKQIQLARPPYLAENHY
AMPDVHAEAGVAVGTVFATERHIVPAGLGGDLGGVAVAFDFPMSLGEVELTRLLA

```



Sat Oct 4 14:14:10 2003

us-10-014-717-1\_copy\_62369\_63628.rge

Page 18

Search completed: October 4, 2003, 03:31:07  
Job time : 4517.41 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:01 ; Search time 8424.03 Seconds  
(without alignments)  
12307.996 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_7610\_11875

Perfect score: 4266  
Sequence: 1 gtcggcgatcgcccatcga.....gggaatcattgccctatga 4266

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	211	4.9	871 29	B2566693 pac82-164
2	188	4.4	753 29	B2561977 pac82-164
3	180.2	4.2	979 29	B2568491 pac82-164
4	165	3.9	1303 29	B2577421 msb2_5402

5	129.4	3.0	865 29	B2566683 pac82-164
6	128.2	3.0	679 29	B2562208 pac82-164
7	126	3.0	1348 29	B2569008 pac82-164
8	123.4	2.9	988 13	BX384927
9	123	2.9	4176 11	BC021544
10	123	2.9	4185 11	BC014631
11	118.8	2.8	1061 13	BQ069153
12	118.2	2.8	1201 13	BX32425
13	114.6	2.7	1798 29	AG171124
14	114.4	2.7	599 29	B2560333 pac82-164
15	113.4	2.7	899 13	BUS28700
16	112.8	2.6	1798 29	AG171124
17	110	2.6	876 29	B2573950
18	108	2.5	1101 29	AG039543
19	107.8	2.5	1196 13	BQ069593
20	107.8	2.5	1198 29	CC190008
21	107	2.5	1438 29	B2557931
22	106.6	2.5	1785 29	CC219595
23	105.6	2.5	1281 12	BG852363
24	105.6	2.5	1695 29	CC290874
25	105	2.5	577 12	BM646821
26	105	2.5	577 12	BM646821
27	104.6	2.5	1299 29	AG039481
28	104.4	2.4	1610 29	B2569386
29	104.2	2.4	690 10	BB865539
30	104	2.4	692 12	BM652119
31	103.6	2.4	1189 29	AG032118
32	103.4	2.4	968 13	BUS57094
33	102.8	2.4	1090 29	B2562999
34	102.6	2.4	625 12	BM091001
35	102.2	2.4	766 10	BE299394
36	102.2	2.4	1651 12	BG809816
37	102	2.4	600 29	B2560280
38	102	2.4	928 13	BQ957144
39	101.6	2.4	699 10	BF250997
40	101.4	2.4	1542 29	AG032943
41	101	2.4	598 29	B2560289
42	100	2.3	704 28	AQ449661
43	99.8	2.3	731 10	BG698437
44	99.8	2.3	1242 12	BM911414
45	99.8	2.3	1350 12	BM450237

## ALIGNMENTS

RESULT 1  
B2566693/c 871 bp DNA linear GSS 17-DEC-2002  
LOCUS pac82-164\_6525.xl pac82-164 Pseudomonas aeruginosa genomic clone  
DEFINITION pac82-164\_6525, genomic survey sequence.  
ACCESSION B2566693  
VERSION B2566693.1 GI:27196732  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

REFERENCE  
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: Shotgun.  
LcLocation/Qualifiers

## FEATURES

```

source
1. 871
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164.6525"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."

BASE COUNT      165 a      294 c      269 g      140 t      3 others
ORIGIN

Query Match      4.9%; Score 211; DB 29; Length 871;
Best Local Similarity 57.6%; Pred. No. 2.5e-27;
Matches 396; Conservative 0; Mismatches 290; Indels 1; Gaps 1;

QY 221 CTTTCCTGAGCAGCAGCCTGCTTTCGACGCTCTTCTTCGACATCTCGCTCGGAG 280
DB 768 CTTTCCTGATGAGGAGGAGGCTTCGAGC-GCGTTCTTCGGGATATCGCTTCGAGG 710
QY 281 CGCTGCGATGAGACCTTCGACATCGACTCTTCGAGGTGTGCGAGGCGCTGAGAG 340
DB 709 CTGCGGAATGAAACCCCGACAGCGCTGCTGAGAGTCTGCGGAATGATCATGGAAC 650
QY 341 AGCGCGGATCGCTCCATCGGCGCTGCTGATCGGAACGGGAGTTCATCGGGATG 400
DB 649 GGGCGGGGCTGCTGCTGTAACCGCTGAAGGTTCCCGCATCGGCTGACGTGGCATC 590
QY 401 GCCCGTCCGATATGAGCGCGCTGCGCGCAAGCGAGCGCTCGGACGATCGAGCTC 460
DB 589 TGGGATGATTAATCTTACCTTACCTGCAATGCGCGAGGCGGATATCCGGGATCAATCAT 530
QY 461 ATGCGCGCTGCGAGCATGCCAGCGTGTGAGCGGCGGATCTGTATGCTTCGCGGC 520
DB 529 TCTATGCGGCTGCGCAAGGATTCAGTTCGCGCGCGCGCATGCGCTATCACTTGGCG 470
QY 521 TCGAGAGGCGGTGTGTCGCGGATGAGATGAGGCTATTCGCTGCTGAGTGGCGCTTATC 580
DB 469 TCGAGGCTCGGAGATGCGGATGAGCAACGCTCTCTTCTGCTGAGTGGGATTCAC 410
QY 581 TCGCGCTGTCAGACCTTCGCTCGCGGAAATGCTTCAAGCGCGCTGCTGAGGATTCG 640
DB 409 TGGGTTGTGCGCCTTCGAGAAATCGGAGGCGGACCTGCGATAGCGCTGGCGTCAATC 350
QY 641 TATGTTGTGCGCGAGCACTTCGTGTGCTCTCGAAGACCGGCGCTGCGCAAGGAGC 700
DB 349 TATGCTGTGCGCGGAGCTGACTATCTCATGAGCGCAATCAAGGCGATTTGCGCGAGCG 290
QY 701 GTGCGTGAAGGATTTTGGGCGGAGCGGATGGGTTTGGAGAGGCGGAAGGTTGCGCG 760
DB 289 GACGCTGCGGGGCTTCGAGCGCGCGCGATGAGATCGTCCGCGCGAAGGAGGTTGCGG 230
QY 761 TCGGTGCTTCAAGCGGCTCAGTGAAGCGCGCGCGGACGCGATCGGATATTTGGCGTGA 820
DB 229 CGTTGTGCTCAAGCGTCTGAGAGAGCTCGAGCGGACCGGACCGGATCTGGGGGTGA 170
QY 821 TTGAGGATTCGCGATCAATCAAGAGGTGCGAGAGCGGCTGATCGTTCGGAAGGGA 880
DB 169 TTGTTGCTGCGGATCAACAGAGCGGCGAGCGCGCGGACCGGATCGGATCGGATCGCA 110
QY 881 GCTCCCAAGAAATCGGTGAACGCGG 907
DB 109 ATGCCAAGCGCGCTGTGGGCTGCGAG 83

RESULT 2
B2561977      753 bp      DNA      linear      GSS 17-DEC-2002
LOCUS
DEFINITION
pac2-164_3530.x1 pac2-164 Pseudomonas aeruginosa genomic clone
ACCESSION
B2561977
VERSION
B2561977.1 GI:27183081
KEYWORDS
GSS.

```

```

SOURCE
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 753)
AUTHORS
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE
Whole-genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol., (2002) In press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1. 753
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164.3530"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."

BASE COUNT      127 a      250 c      243 g      133 t
ORIGIN

Query Match      4.4%; Score 188; DB 29; Length 753;
Best Local Similarity 60.3%; Pred. No. 2.7e-23;
Matches 311; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 633 GGTATCGCTGATGTTGTGTCGCGAGCACCTGCTGTGCTTCGAGAACCCGCGCTGCG 692
DB 95 GCTTCCCTGATCCTTTGCGCCTATGTTGTGCTGTGCTGACGAAATTCGCGCGGTGCG 154
QY 693 CAGGAGCGGTGCTGCAAGGATTTTCGCGGAGCGCATGAGTTTGGACGAGCGGAAG 752
DB 155 CAGCATGAGGCGTTTCAAGACCTTTCAGCGCGCGCGCATGAGCTGAGCGGAGAGGG 214
QY 753 GTGCGCGTGTGCTGCTCAAGCGGCTGAGTGAAGCGCGCGGAGCGGATCGGATATT 812
DB 215 CTGCTGTGCTGCTGTTCTCAAGCGCTGCGCGAGCGCTGACCGGATTCGCTCT 274
QY 813 GCGGTGATTCGAGGATCCGCGATCAATCAGACGCTGCGAGCGGCTTGAACCGTCC 872
DB 275 CCGGTTGATCCAGGTTAGCGGCTTCAACAGACGCGCTTCAGCGAGCGGAATCAAGGTTCC 334
QY 873 GAAAGGAGCTTCCAAAGAAATGTGCTGAAGCGGCTTCGCGGACGAGGCTGCGCGC 932
DB 335 GAAATTCACCGCCGAGAGCGGTGATGCGCGCTCTTGGCGCAATGCGGAGCGCAAGGC 394
QY 933 GTTTCGTTGCTTATGTCGAGGACACGAGCGGCGGACGAGCTTGTGTAACCCATGA 992
DB 395 GAGCGAGTCACTACATGAGGACATGATGACGCGACTCGCTGGGCGGAGCGGATGGA 454
QY 993 AATCCAGCTCTGATGCGGTATAGCGGCTCGGCGGAGATGTGCGACGCGCTGAT 1052
DB 455 ACTGCGCGGCTGATGCGGTGCTGCGGCGGAGCGCGGCTGACAGGCGGCTGCTGAT 514
QY 1053 CGGTCGCTGAAGACCAACTTGGGCACTTCGATGATGCGTGGGAGATCACTGGCTGCT 1112
DB 515 GGGATCGATTAAGGCAACATCGGCACTCGAGCGCATGCGCGCGCTGCGGAGCTGCG 574
QY 1113 GAAAGTGTCTTGTCTTCAGACGAGGAGATTCC 1148
DB 575 GAAATCTGCTGCTGCTTGAACAGCAACGATTGCC 610

RESULT 3

```



B2568491/c 979 bp DNA linear GSS 17-DEC-2002  
 LOCUS B2568491  
 DEFINITION pacs2-164\_7558.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
 pacs2-164\_7558, genomic survey sequence.  
 ACCESSION B2568491  
 VERSION B2568491.1 GI:27201527  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 979)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.  
 FEATURES  
 source 1..979  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164\_7558"  
 /note="Clinical isolate 2-164 whole genomic shotgun  
 library."  
 BASE COUNT 187 a 324 c 299 g 164 t 5 others  
 ORIGIN  
 Query Match 4.2% Score 180.2; DB 29; Length 979;  
 Best Local Similarity 55.1%; Pred. No. 6,5e-22;  
 Matches 372; Conservative 0; Mismatches 297; Indels 6; Gaps 1;  
 Oy 186 CGATGCCCGGGGAGACGCCCGCTTACGCGCATCTTCTCGACGACGTAGCTGCTT 245  
 Db 753 CCAAAACCGGTCACACACCGCGGGCCATGCAAGCTTGTCAGCGGTGAAGAGGGTT 694  
 Oy 246 CGAGCCCTCTTCTTGGCATCTCGCTCTCGAAGCGCTCGGATGACCTGCACATCG 305  
 Db 693 CACCAAGNACTTTTCCGATCTCCGACAGGNAACGGCAAGCATGATCGGCAACAG 634  
 Oy 306 ACTCTGTGAGGTGTCTGAGAGCGCTGAGAAACCGCGATCGCTCCATCGGCCCT 365  
 Db 633 CTTGTTCTGATGTGTGCTGAGGCGCTGAGAGGGGAGGTATTCGCGCATGACTT 574  
 Oy 366 CGTCGTAACGGAACGGAAGTGTTCATCGGATCGGCGCGCTCGAATATGAGCGCGCT 425  
 Db 573 GCGGGAAGCCAGTCGTGTTTCTGTGGGCGCAATGCCAGCTAGAGGACGTTT 514  
 Oy 426 GCCCAAGCAGCGGCTCGCAGAGATGACGCTCATGGCGGGCTGGGAGCAGTCCAG 485  
 Db 513 GCTGG-----CTGCGCGCGCGGATGACCGCAACTACGGTACCGGAGGCGCTTCTC 460  
 Oy 486 CGTCGAGCGGGCGAATCTGTATGCCCTGCGGGCTGGAGGGCGGTGTGCGGGTGA 545  
 Db 459 GGGCATTTCCGGCGCATCTGCAATTTCTTGCCCTGGCGGTCTCTCCCTGACCTGA 400  
 Oy 546 TACGCGCTATTCGTCCTGCTGTCGCGCTTCATCTGACCTGTCAGAGCTTGCCTCGG 605  
 Db 399 CACGCGGTGCTGTATCGCTGACGCGCTGTGCACTTGGCTGCAACAGCTGTAGCGG 340  
 Oy 606 GGAATGTCACAGCGCTGCTGCTGTGGGGTATCGCTGATTTGTGCGGAGCACCCTGT 665  
 Db 339 GGAGTGCAGATGTCTATGTCGCGGGGTGAACGTACATCGCATCTCGGCAATCTTCA 280

Oy 666 GTGCTCTCGAAGACCGGCGCTGCGCAGGACGTCGCTGCAAGGCAATTTGCGCGGA 725  
 Db 279 ATCATGGGCGCATCCGCGGTGCTATCGGCGGACGACCTGCAAGACTTTCAGACATCG 220  
 Oy 726 GCGCATGGGTTTGGACGAGCGGAAGGTGCGCGCTGCTGCTCTCAAGCGGCTCAGTCG 785  
 Db 219 TGCCGATGGCTACCGCGCGCGGAGGTGCGGTGTGTCTCAAGCGGCGGAGCG 160  
 Oy 786 AGCCCGCGCGGACCGCATGCGGATATGCGCGGTATTCGAGATCCGATCAATCAGCA 845  
 Db 159 GCGCGGCGGACGCGCAGCGGGTCAATCGCTTCTGCGGCTCGGCGGTCAACATCA 100  
 Oy 846 CGTGCAGACGACCGG 860  
 Db 99 CGGCGCTGCGCGGG 85  
 RESULT 4  
 B2577421 1303 bp DNA linear GSS 17-DEC-2002  
 LOCUS B2577421  
 DEFINITION meh2\_5402.x1 meh Pseudomonas aeruginosa genomic clone meh2\_5402,  
 genomic survey sequence.  
 ACCESSION B2577421  
 VERSION B2577421.1 GI:27212482  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 1303)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.  
 FEATURES  
 source 1..1303  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="meh2\_5402"  
 /note="Environmental isolate. Whole genomic shotgun  
 library."  
 BASE COUNT 278 a 362 c 359 g 295 t 9 others  
 ORIGIN  
 Query Match 3.9% Score 165; DB 29; Length 1303;  
 Best Local Similarity 57.9%; Pred. No. 3.1e-19;  
 Matches 307; Conservative 0; Mismatches 220; Indels 3; Gaps 1;  
 Oy 285 GCGATGACCCCTGACATGACCTTGTGAGAGGTGCTGGAAGCGCTGAGAAACG 344  
 Db 94 GGGCATGACCCCGCAACGCAATCGCGCTGAAAGTTCCTGGAAAGCTGAGACATGC 153  
 Oy 345 GCGATGCTTCATCGGCGCTGTCGATGCGAAACGGAAGTTCATCGGAGATCGGCC 404  
 Db 154 GGTATGACCCGTTTCCCTCGATGCGCAAGCTGTGCTTATGAGGGGCGCAGTAC 213  
 Oy 405 GTCCGATATAGAGCGCGCTGCGCAAGCAACGCGCTCGCAGAGATGACAGCTTCATCG 464  
 Db 214 CAAGACTATATAGCGCTGCGCAGCAACTCGGAGAGGTGCTGACGTCAATGCGCA--- 270

OY		465	CGGGCTGGGAGACGAATGCCCAAGCGTGGAGCCGAGCCGAATCTGTATAGCCTTCGGGCTGGG	524
Db		271	CCAATTCTAACGGCAGGACCAGCTTTATGCGCCAGATATGCTACACCTTGNGCGGACA	330
OY		525	AGGAGCGGTGTGTGGCGGTGGATACGAGCCATTTCGTCTTCGTGTGGCCGTTTCATCTGGC	584
Db		331	AGGCCCGGAGTAGATGATGACACCTCTGTCGTCTTTCGTGTGGAGCGCTGACCAAGC	390
OY		565	CTGTCAAGACTTGGCGCTCCGGGAAATGCTCACAGGCCCTTGGCTGGTGGGGTATCCGTAT	644
Db		391	CTGCCAAGAGCCTTGGCCGACGGGGAAGAACAACCTGACGCTGGCGCGAGTGAACCTGAT	450
OY		645	GTTTGGCCGAGACACCTCTGTGTGCTCCGAAGACCCGGGCGCTGGACGAGGACGATG	704
Db		451	CCTTTGCGCCCTATGGTTTCGTCTGTCTGTACGAAATTTGGGGCGGTGCACAGCATGGGCG	510
OY		705	CTGCAAGGCAATTTTCGGCGGAGCCGATGGATTTCGACAGAGCGAAGGCTGGCCCTGT	764
Db		511	TTCGAAGACCTTTCGACGCGCGCGNCGACGGCTATGGCCCTGAGAGAGGCGTGGCTGTGT	570
OY		765	GGTCTCTCAAGCCGCTCAGTGGAGCCCGCGCGGACGCGCATCGGATATTGG	814
Db		571	TGTTCTCAAGCCGCTTGGCGGACGCGGACGCGTACCGTATCCGTGCTCGG	620
RESULT 5				
BZ566683				
LOCUS	BZ566683	865 bp	DNA	linear GSS 17-DEC-2007
DEFINITION	pacs2-164_6520..x1 pacs2-164 Pseudomonas aeruginosa genomic clone			
ACCESSION	pacs2-164_6520, genomic survey sequence.			
VERSION	BZ566683.1 GI:27196700			
KEYWORDS	GSS.			
SOURCE	Pseudomonas aeruginosa			
ORGANISM	Pseudomonas aeruginosa			
	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
	Pseudomonadaceae; Pseudomonas.			
REFERENCE	1 (bases 1 to 865)			
AUTHORS	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.			
TITLE	Whole-genome-sequence variation among multiple isolates of Pseudomonas aeruginosa library			
JOURNAL	J. Bacteriol., (2002) In press			
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: ckraymond@u.washington.edu Class: shotgun. Location/Qualifiers			
FEATURES	source	1..865		
	/organism="Pseudomonas aeruginosa"			
	/mol_type="genomic DNA"			
	/strain="2-164"			
	/db_xref="taxon:287"			
	/clone="pacs2-164_6520"			
	/clone_lhb="pacs2-164"			
	/note="clinical isolate 2-164 whole genomic shotgun library."			
BASE COUNT	142 a	263 c	298 g	161 t
ORIGIN				1 others
Query Match		3.0%	Score 129.4; DB 29; Length 865;	
Best Local Similarity		53.5%; Pred. No. 5.2e-13;		
Matches	296; Conservative	0; Mismatches 251; Indels	6; Gaps	1
OY		756	CGCCGTCTGTGTCTCAAGCGGCTCAGTGGAGCCCGCGGACGCGCATCGGATATTGGC	815
Db		45	CGAGTTCGACGGTATTCGATAAGACTTGATATCGAATTCCTCGACGCGGACCGGGTCATCGC	104
OY		816	GGTGATTGAGGATCCGCGATCAATCAGACGGGTGCGAGCGGCTGACCGGTGCGGA	875

Db	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	
Db	B2562208/c	B2562208	pac82-164_3644	GI:27183746	GSS.	Pseudomonas aeruginosa	Spencer, D. H., Raymond, C. K., Smith, E. E., Sims, E. E., Hastings, M., Burns, J. L., Kaul, R. and Olsen, M. V.	Whole-Genome-Sequence Variation among multiple isolates of Pseudomonas aeruginosa library	J. Bacteriol., (2002) In press	Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymond@u.washington.edu Class: Shotgun.		
Qy	105	CGTTCTGTGGGGCTCGGCGGTCAACATGACGGGGCCCTGGCGGGGGCTCACAGTACCCAA	105	CGTTCTGTGGGGCTCGGCGGTCAACATGACGGGGCCCTGGCGGGGGCTCACAGTACCCAA	105	CGTTCTGTGGGGCTCGGCGGTCAACATGACGGGGCCCTGGCGGGGGCTCACAGTACCCAA	105	CGTTCTGTGGGGCTCGGCGGTCAACATGACGGGGCCCTGGCGGGGGCTCACAGTACCCAA	105	CGTTCTGTGGGGCTCGGCGGTCAACATGACGGGGCCCTGGCGGGGGCTCACAGTACCCAA	105	CGTTCTGTGGGGCTCGGCGGTCAACATGACGGGGCCCTGGCGGGGGCTCACAGTACCCAA
Qy	876	CGGAGACTCCCAAGAAATCGTGTGAAACGGGGCCCTGGCGGACCGAGGCTGCGCGCTC	876	CGGAGACTCCCAAGAAATCGTGTGAAACGGGGCCCTGGCGGACCGAGGCTGCGCGCTC	876	CGGAGACTCCCAAGAAATCGTGTGAAACGGGGCCCTGGCGGACCGAGGCTGCGCGCTC	876	CGGAGACTCCCAAGAAATCGTGTGAAACGGGGCCCTGGCGGACCGAGGCTGCGCGCTC	876	CGGAGACTCCCAAGAAATCGTGTGAAACGGGGCCCTGGCGGACCGAGGCTGCGCGCTC	876	CGGAGACTCCCAAGAAATCGTGTGAAACGGGGCCCTGGCGGACCGAGGCTGCGCGCTC
Db	165	CGGACCGGCCCCAGAGAGGCGTTGATTTCGCGAAGCCCTGGCCAAACCCAGCTGTGCGCGC	165	CGGACCGGCCCCAGAGAGGCGTTGATTTCGCGAAGCCCTGGCCAAACCCAGCTGTGCGCGC	165	CGGACCGGCCCCAGAGAGGCGTTGATTTCGCGAAGCCCTGGCCAAACCCAGCTGTGCGCGC	165	CGGACCGGCCCCAGAGAGGCGTTGATTTCGCGAAGCCCTGGCCAAACCCAGCTGTGCGCGC	165	CGGACCGGCCCCAGAGAGGCGTTGATTTCGCGAAGCCCTGGCCAAACCCAGCTGTGCGCGC	165	CGGACCGGCCCCAGAGAGGCGTTGATTTCGCGAAGCCCTGGCCAAACCCAGCTGTGCGCGC
Qy	936	TTTCGTTGGGCTTATGTTCAGGCGACACCGGACGGGGACGACGCTTGGTACCCCATCGAAAT	936	TTTCGTTGGGCTTATGTTCAGGCGACACCGGACGGGGACGACGCTTGGTACCCCATCGAAAT	936	TTTCGTTGGGCTTATGTTCAGGCGACACCGGACGGGGACGACGCTTGGTACCCCATCGAAAT	936	TTTCGTTGGGCTTATGTTCAGGCGACACCGGACGGGGACGACGCTTGGTACCCCATCGAAAT	936	TTTCGTTGGGCTTATGTTCAGGCGACACCGGACGGGGACGACGCTTGGTACCCCATCGAAAT	936	TTTCGTTGGGCTTATGTTCAGGCGACACCGGACGGGGACGACGCTTGGTACCCCATCGAAAT
Db	225	GCGGATGGGCTATGTTCAGGCGCCCATGGGACCGGACCACTGCTTCGCGATCCGATCGAGTT	225	GCGGATGGGCTATGTTCAGGCGCCCATGGGACCGGACCACTGCTTCGCGATCCGATCGAGTT	225	GCGGATGGGCTATGTTCAGGCGCCCATGGGACCGGACCACTGCTTCGCGATCCGATCGAGTT	225	GCGGATGGGCTATGTTCAGGCGCCCATGGGACCGGACCACTGCTTCGCGATCCGATCGAGTT	225	GCGGATGGGCTATGTTCAGGCGCCCATGGGACCGGACCACTGCTTCGCGATCCGATCGAGTT	225	GCGGATGGGCTATGTTCAGGCGCCCATGGGACCGGACCACTGCTTCGCGATCCGATCGAGTT
Qy	996	CCAAAGCTTCGATGATGCGGATATACCGGCTTCGCGGAGATGTCCGACCGCGCTGTGATCGG	996	CCAAAGCTTCGATGATGCGGATATACCGGCTTCGCGGAGATGTCCGACCGCGCTGTGATCGG	996	CCAAAGCTTCGATGATGCGGATATACCGGCTTCGCGGAGATGTCCGACCGCGCTGTGATCGG	996	CCAAAGCTTCGATGATGCGGATATACCGGCTTCGCGGAGATGTCCGACCGCGCTGTGATCGG	996	CCAAAGCTTCGATGATGCGGATATACCGGCTTCGCGGAGATGTCCGACCGCGCTGTGATCGG	996	CCAAAGCTTCGATGATGCGGATATACCGGCTTCGCGGAGATGTCCGACCGCGCTGTGATCGG
Db	285	GCAAGGCCCT-----GGCCAAAGCGCTATTCAGAGTACTGAGGCGCGGCTTGGGCGGTGC	285	GCAAGGCCCT-----GGCCAAAGCGCTATTCAGAGTACTGAGGCGCGGCTTGGGCGGTGC	285	GCAAGGCCCT-----GGCCAAAGCGCTATTCAGAGTACTGAGGCGCGGCTTGGGCGGTGC	285	GCAAGGCCCT-----GGCCAAAGCGCTATTCAGAGTACTGAGGCGCGGCTTGGGCGGTGC	285	GCAAGGCCCT-----GGCCAAAGCGCTATTCAGAGTACTGAGGCGCGGCTTGGGCGGTGC	285	GCAAGGCCCT-----GGCCAAAGCGCTATTCAGAGTACTGAGGCGCGGCTTGGGCGGTGC
Qy	1056	GTCGGTGAAGCAACAACCTTGGCCATCTCTGATATGCGTGGGGATCACTGGGGCTGTGAA	1056	GTCGGTGAAGCAACAACCTTGGCCATCTCTGATATGCGTGGGGATCACTGGGGCTGTGAA	1056	GTCGGTGAAGCAACAACCTTGGCCATCTCTGATATGCGTGGGGATCACTGGGGCTGTGAA	1056	GTCGGTGAAGCAACAACCTTGGCCATCTCTGATATGCGTGGGGATCACTGGGGCTGTGAA	1056	GTCGGTGAAGCAACAACCTTGGCCATCTCTGATATGCGTGGGGATCACTGGGGCTGTGAA	1056	GTCGGTGAAGCAACAACCTTGGCCATCTCTGATATGCGTGGGGATCACTGGGGCTGTGAA
Db	339	ATCGGTGAAGATTAACGTTCGACATCTGGAGAGCCGCGCCGCTATTCGTTGTGATCA	339	ATCGGTGAAGATTAACGTTCGACATCTGGAGAGCCGCGCCGCTATTCGTTGTGATCA	339	ATCGGTGAAGATTAACGTTCGACATCTGGAGAGCCGCGCCGCTATTCGTTGTGATCA	339	ATCGGTGAAGATTAACGTTCGACATCTGGAGAGCCGCGCCGCTATTCGTTGTGATCA	339	ATCGGTGAAGATTAACGTTCGACATCTGGAGAGCCGCGCCGCTATTCGTTGTGATCA	339	ATCGGTGAAGATTAACGTTCGACATCTGGAGAGCCGCGCCGCTATTCGTTGTGATCA
Qy	1116	GATGCTGTTCCTTCCTTCAGCAAGGGGAGATTCTTCGCGCACTCCACGCGGAGGGCGCTGAA	1116	GATGCTGTTCCTTCCTTCAGCAAGGGGAGATTCTTCGCGCACTCCACGCGGAGGGCGCTGAA	1116	GATGCTGTTCCTTCCTTCAGCAAGGGGAGATTCTTCGCGCACTCCACGCGGAGGGCGCTGAA	1116	GATGCTGTTCCTTCCTTCAGCAAGGGGAGATTCTTCGCGCACTCCACGCGGAGGGCGCTGAA	1116	GATGCTGTTCCTTCCTTCAGCAAGGGGAGATTCTTCGCGCACTCCACGCGGAGGGCGCTGAA	1116	GATGCTGTTCCTTCCTTCAGCAAGGGGAGATTCTTCGCGCACTCCACGCGGAGGGCGCTGAA
Db	399	GAGCTGCTGT	399	GAGCTGCTGT	399	GAGCTGCTGT	399	GAGCTGCTGT	399	GAGCTGCTGT	399	GAGCTGCTGT
Qy	1176	CCCCCGAATCATAGGAGGTGATCTTTCGCGCTGACCGCTGACGCGCCCGGACACCGTGGCC	1176	CCCCCGAATCATAGGAGGTGATCTTTCGCGCTGACCGCTGACGCGCCCGGACACCGTGGCC	1176	CCCCCGAATCATAGGAGGTGATCTTTCGCGCTGACCGCTGACGCGCCCGGACACCGTGGCC	1176	CCCCCGAATCATAGGAGGTGATCTTTCGCGCTGACCGCTGACGCGCCCGGACACCGTGGCC	1176	CCCCCGAATCATAGGAGGTGATCTTTCGCGCTGACCGCTGACGCGCCCGGACACCGTGGCC	1176	CCCCCGAATCATAGGAGGTGATCTTTCGCGCTGACCGCTGACGCGCCCGGACACCGTGGCC
Db	459	CAGCCGCTTCAGCTGTGACGAGCAAGAGTCTCCAGTGTGCGAGGCAATCAGCCCTGGGA	459	CAGCCGCTTCAGCTGTGAC								

## ORIGIN

Query Match 3.0%; Score 128.2; DB 29; Length 679;  
Best Local Similarity 59.3%; Pred. No. 8.3e-13;  
Matches 239; Conservative 0; Mismatches 158; Indels 6; Gaps 1;

QY 1623 GCGCTTTCTTTACCGGACAGGGGGCGCAGACGCTGGGCGATGGCCCTGGGCTGTACGA 1682  
DB 515 GCGCTGCTGTATCCCGGCGCAGGGGCTGCCAGATCCGGGAAATGGCCAGGACGCTGTACGA 456  
QY 1683 TGTATGTCCTGCTCCCGAGGCGCTTGAACCTGTGCGTGGAGGCTTTCAACAGAGACT 1742  
DB 455 CAGCAGAGCTGCGCTTCCGTGAGGCAATGAGACAGTGTCTGTGGCTGGCGGAGATGG 396  
QY 1743 CGACCGGCGGCTCCCGAGAGTGTATGGGCGCAACCGGCGACGCTGCAAGCGCGCTGCT 1802  
DB 395 TGTGAGTGTCTTCCGCTTGT-----CGAGTCAAGCGCGGAGGAGGAGGCTTGT 342  
QY 1803 CGACCAAGACGCTTACCCAGCGGCGCTGTTCACCTTGAATATGCGCTGCGCGCT 1862  
DB 341 GCGCGAGACAGCTATGCGCAGCGCGGCAATTTCCCGTGAATACGCGCTGCGCGCT 282  
QY 1863 GTGCGGCTGTGGGCTGTAGAGCCGAGTGTGCTCCCGCCATACATCGTGTAGCTGT 1922  
DB 281 GTGATGAGCTGTGGGCGGTCTCCCGGCGCGCTGTTCAGTTCAGTTCGTGAATATGG 222  
QY 1923 GAGTGTGCTGTGGGCGGCTGTCTCGCTTGAAGAGCGGCTTCTGTGGCTGTGGCG 1982  
DB 221 GCGCGCTGTATCGCGGGGCGCTTCAAGCTGAGGAGCGCGCTTGAAGATGTGAGGCGCG 162  
QY 1983 CGGCGGCTGTATGACAGGCGCTGCGGCGGCGGCGGCGATGTGTG 2025  
DB 161 CGAGCGCTGGGCGCAGAACTGATGACGCGCGGAGGGGTGATG 119

RESULT 7  
B2569008/c 1348 bp DNA linear GSS 17-DEC-2002

LOCUS pacs2-164.8195.x1 pacs2-164 Pseudomonas aeruginosa genomic clone  
DEFINITION B2569008.1 GI:27202907

ACCESSION B2569008.1  
VERSION B2569008.1  
KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa

REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

AUTHORS 1 (bases 1 to 1348)  
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R., and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library

JOURNAL J. Bacteriol., (2002) In press

COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: cgraymond@u.washington.edu  
Class: shotgun.

## FEATURES

source

1..1348  
Location/Qualifiers

/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"

/db\_xref="taxon:287"  
/clone="pacs2-164.8195"  
/note="Clinical isolate 2-164 Whole genomic shotgun  
library."

BASE COUNT 319 a 389 c 348 g 274 t 18 others  
ORIGIN

Query Match 3.0%; Score 126; DB 29; Length 1348;  
Best Local Similarity 51.6%; Pred. No. 2.1e-12;  
Matches 299; Conservative 0; Mismatches 279; Indels 1; Gaps 1;

QY 129 CGTGGGCGAGTCCCGCCGCAACGCTGGAGTACAGACGCTGTATCCCGACCCCGA 188  
DB 674 CGTCGCGAGTCTCCCGCGGCGCTGTGAATGGCG-AGCGATTTCACGACCGCTGTGCGA 616  
QY 189 TGCCCCGGGGAAGCGCCCGTTACGCGCGCATCTTCTCGAGCGACGATAGCTGTTGA 248  
DB 615 GCGTCCGAGTCACTATGCGNNNGAGAGGCGCTTCTTATATGAGTGAACGCTTGA 556  
QY 249 GCGCTCTCTTTCTGGGCACTCGCTCGGAAAGCGCTGCGAGTACCTGCAATGCACT 308  
DB 555 GCGCGGTTTCTTCCGGAATCCGCTTCTGAGGCTTCGGAATGACCTTCGAGCGGCT 496  
QY 309 CTGCTGAGAGTGTGCTGGAGGCGCTGAGAGAGCGCGCATCGCTCATGCGCGCTGT 368  
DB 495 GCTCTGAGAGTGTGCTGGAGTCAATGAGAAAGCGGCGGCGCTGCTGAAACGCTGAA 436  
QY 369 CGGTACGGAAGCGGAGTTCATCGGATAGCGCCCGTCCGAATATGAGCGCGCTGCC 428  
DB 435 AGGTTCGCGCACTGCGGTCTAGTGGCATCTGGGCGATGATTAACCTTACCTGCAATGC 376  
QY 429 GCAAGCGACGCGCGTCCGAGAGTCAAGCTCATGCGGCGTGGGAGAGATGCCAGCGT 488  
DB 375 GCGGAGGCGCGTATCCCGGAGATCAATTCATTCATGCGGCTGCGAAGATTCAGCTT 316  
QY 489 CGAGCGGCGCGGAATCTGTATGCGCTGCGGCGTGGAGGCGCTGTGTGCTGCGGTGATAC 548  
DB 315 GCGCGCGGCGCGATCGGCTATCACTTGGCGGTGACACGCTGCGGCGAGTGGGTGAGACAC 256  
QY 549 GCGCTATTCGCTCTGCTGCGTGGCGGCTTCACTGCGCTGTCAGAGCTTGGCGTCCGCGGA 608  
DB 255 GCGCTGCTCTCTTCTGCTGCGTGGCGGATTCACCTGAGTGTGCGCTTGCAGAAATGCGGA 196  
QY 609 ATGCTCCAGCGCGCTGCTGCTGCGGTATCGCTGATGTGTGCGCGAGACACCTGCTGTG 668  
DB 195 GCGCGAGCTGCGCATAGCGGCTGCGCTCAATCTGATGCTGTGCGCGAGCTGACTATCTTA 136  
QY 669 GCTCTGAGAGACCGGCGGCTGGCGGAGGAGCGTGTGCTG 707  
DB 135 CATAGCGCAGATCAAGGCGATTTGCGCGAGCGGAGCGCTG 97

RESULT 8  
BX384927 988 bp mRNA linear EST 08-MAY-2003

LOCUS BX384927 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0D0004YA22 5-PRIME, mRNA sequence.

ACCESSION BX384927  
VERSION BX384927.1  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 988)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

AUTHORS Full-length cDNA libraries and normalization  
JOURNAL Unpublished

COMMENT Genoscope  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr; web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 8574.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D0004B11Q1pkcluster=8574.r. Contact :  
Peng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600

FEATURES Faraday Avenue Genoscope sequence ID : CS0DL004B11QPI.  
Location/Qualifiers

1..988  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL004YA22"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_11b="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 184 a 334 c 304 g 166 t  
ORIGIN

Query Match 2.9%; Score 123.4; DB 13; Length 988;  
Best Local Similarity 53.0%; Pred. No. 5.9e-12;  
Matches 290; Conservative 0; Mismatches 251; Indels 6; Gaps 1;

226 CTGAGGAGGAGTGGCTGCTTTCAGAGCGCTCTTTCGAGATCTCGCTCGGAGAGCGCTG 285  
328 CTGAGGAGGAGTGGCTGCTTTCAGAGCGCTCTTTCGAGATCTCGCTCGGAGAGCGCTG 387  
286 CGGATGAGACCTGACATGACTCTTCTGAGAGTGTCTGGAGGCGCTGGAGAGCGCC 345  
388 AGGATGAGACCTGACATGACTCTTCTGAGAGTGTCTGGAGGCGCTGGAGAGCGCC 447  
346 GGGATGAGACCTGACATGACTCTTCTGAGAGTGTCTGGAGGCGCTGGAGAGCGCC 405  
448 GGGATGAGACCTGACATGACTCTTCTGAGAGTGTCTGGAGGCGCTGGAGAGCGCC 501  
406 TCCGATATGAGAGCGCGCTGCGCAGAGCGCGCTGCGCAGAGATCGATCGATGCGC 465  
502 AGCGGCTGAGAGCGCGCTGCGCAGAGCGCGCTGCGCAGAGATCGATCGATGCGC 561  
466 GGGCTGAGAGCGCGCTGCGCAGAGCGCGCTGCGCAGAGATCGATCGATGCGC 525  
562 ATGAGGAGGAGCGCGCTGCGCAGAGCGCGCTGCGCAGAGATCGATCGATGCGC 621  
526 GGGCTGAGAGCGCGCTGCGCAGAGCGCGCTGCGCAGAGATCGATCGATGCGC 585  
622 GGGCTGAGAGCGCGCTGCGCAGAGCGCGCTGCGCAGAGATCGATCGATGCGC 681  
586 TGTCAAGAGCTGCGCTCGGAGAGTGTCTCAAGCGCGCTGCGTGGAGGATCGATGCG 645  
682 TACCAAGAGCTGCGCTCGGAGAGTGTCTCAAGCGCGCTGCGTGGAGGATCGATGCG 741  
646 TGTGAGGAGAGCGCTCGGAGAGTGTCTCAAGCGCGCTGCGTGGAGGATCGATGCG 705  
742 CTGAAGAGAGCGCTCGGAGAGTGTCTCAAGCGCGCTGCGTGGAGGATCGATGCG 801  
706 TGAAGAGAGCGCTCGGAGAGTGTCTCAAGCGCGCTGCGTGGAGGATCGATGCG 765  
802 TGAAGAGAGCGCTCGGAGAGTGTCTCAAGCGCGCTGCGTGGAGGATCGATGCG 861  
766 GTCTCTCA 772  
862 CTGCTGA 868

RESULT 9  
LOCUS BC021544 4176 bp mRNA linear HTC 20-MAY-2002  
DEFINITION Homo sapiens, clone IMAGE:3355480, mRNA.  
ACCESSION BC021544  
VERSION BC021544.1 GI:20987241  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Carnivora; Hominiidae; Homo.  
1 (bases 1 to 4176)

AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Ffeli, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeary, Steven Neese, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stolt, Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IPAL Plate: 22 Row: P Column: 6  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis  
This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers

1..4176  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3355480"  
/issue\_type="Eye, retinoblastoma"  
/clone\_11b="NIH MGC\_16"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
BASE COUNT 702 a 1419 c 1319 g 736 t  
ORIGIN

Query Match 2.9%; Score 123; DB 11; Length 4176;  
Best Local Similarity 50.2%; Pred. No. 7.8e-12;  
Matches 340; Conservative 0; Mismatches 325; Indels 12; Gaps 1;

3128 TTGCTACCAAGTCAACAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3187  
1584 TGGCCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1643  
3188 CGATGAGAGCG 3247  
1644 TGAAGAGAGCG 1703  
3248 TAGCCCGCTTCAAGCG 3307  
1704 GTGGCGGAGTGGCG 1763  
3308 TGGCGCGGTGCAAGTGGCG 3367  
1764 AGGCGCGCATCGCCATCGCCCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1823  
3368 CCGAGAAAGCG 3415  
1824 CTGAGAGCG 1883  
3416 ATTCCCGCTCGAAGCGGTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3475  
1884 ACTCCGGGAGACATCTTTCAGAGCATGTGCTGTGCAACAGCGCGCGCGGAGGCGCTTG 1943

QY 3476 ACGTGTCTCACTCGCTCTCGGCGAGCTGATGACAGATTTCATCTCCGCAT 3535  
 Db 1944 ACCTGTCTTGAATCTCTGGCGGAGAGAGCTGACAGCGGTGAGTCTTGCTA 2003  
 QY 3536 CGCAGCGCGGCTTGTGAGCTCGGCAAGCGGACTGTTACGGGATACCAAGCTCGGCG 3595  
 Db 2004 CGCAGCGGCTCTCTGGAATTTGGCAATTGCACTTTCAGAACACCGCGCTCGGCA 2063  
 QY 3596 TGGCGCGGCTCTCGGCAATCTCTCTCTGCTGTGGATCTCCGGGGATGATGCTCG 3655  
 Db 2064 TGGCTATCTTCTTAAGAGCTGATTCACGCGGGTCTTACTGATGCTGTTTCAACG 2123  
 QY 3656 AGCGCGCGCGCGGCTCGCTCGCTTGGAGAGCTCTCGGCTGATCGCGGAGCG 3715  
 Db 2124 AGACAGAGTGTGACTGCGGAGAGGTGTGGCGCTTGTTCAGACCGGATCCGGATGCGG 2183  
 QY 3716 TGTTCACCCCTCCCGCCATCGGACGCTCCGATCGCGGTGTGCGCGATGCTTCGGA 3775  
 Db 2184 TGTACGCGCCCTCAAGTGCACGAGTGTTCATGAGGCGCCAGGTGAGAGACGCTTCGCT 2243  
 QY 3776 GCATGGCGCAGCGCCAG 3792  
 Db 2244 ACATGGCCCAAGGAG 2260

## RESULT 10

BC014631 4185 bp mRNA linear HTC 12-SEP-2002  
 LOCUS BC014631 Homo sapiens, fatty acid synthase, clone IMAGE:3351370, mRNA.  
 DEFINITION BC014631  
 ACCESSION BC014631.1 GI:17939526  
 VERSION HTC.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 4185)  
 STRAUSBERG, R.  
 DIRECT SUBMISSION  
 Submitted (24-SEP-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Institute for Systems Biology  
 http://www.systembiology.org  
 contact: amadan@systembiology.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kesteman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAL Plate: 11 Row: m Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis  
 This clone has the following problem: frame shifted.

## FEATURES

SOURCE  
 1. 4185  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3351370"  
 /rname="EYE, retinoblastoma"  
 /clone\_id="NIH\_MGC\_16"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"

BASE COUNT 704 a 1421 c 1324 g 736 t  
 ORIGIN

Query Match 2.9%; Score 123; DB 11; Length 4185;  
 Best Local Similarity 50.2%; Pred. No. 7.8e-12;  
 Matches 340; Conservative 0; Mismatches 35; Indels 12; Gaps 1;

QY 3128 TTGCTACCCACAGTCAACAGCTCGGCTCGCTGTGCTGCTCGCTCAGCGCTTCG 3187  
 Db 1593 TGGCCACTCTGTCTCTCTGTCTCAACCGAGCTTCTCTGAGATGTCTTCAACTGAGAGC 1652  
 QY 3188 CGATGAGGCGGCGCCATGCGCTCGCTGCTTCACTGACGAGATGATGCTGACAGAA 3247  
 Db 1653 TGAAGAGGCGGCTCGCTGCTGCTGCTTCAACAGACAGCGCTTCACTGAGCTGTGCTG 1712  
 QY 3248 TAGCCGCCCTTCAACCGCGGAGAGCGGTGCTGATTCATTCGCGGCAACCGCGGCTGCTG 3307  
 Db 1713 GTGGCGGAGTGGCGCTCGCGGAGAGCGTGTCTCACTGACAGCTGAGGCTCGGCGCTGAGC 1772  
 QY 3308 TCGCGCGGTGACAGTGGGCGCAGCAGTGGAGCGGAGTTCATGACAGCGCGGAGAGAG 3367  
 Db 1773 AGCGGCGCATGCGCATGCGCTCTCACTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 1832  
 QY 3368 CCGAGAAACGCGCTTACCTGAGTGTGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 3415  
 Db 1833 CTGAGAAAGCGGCGTACCTCAAGGCGCAGGTTCCCGCAGCTGACAGCAGCAGCTTCCGCA 1892  
 QY 3416 ATTCCGCGCTGAGACCGGTTGTGCTGCGCAGCGGCGGCGGAGAGAGAGAGAGAGAGAG 3475  
 Db 1893 ACTCCCGGAGACATCTCTGAGAGAGATGTGCTGTGGACACACGCGGAGAGAGAGAGAGAG 1952  
 QY 3476 ACGTGTCTCACTCGCTCTCGGCGAGCTGATGACAGATTTCATCTCCGCAT 3535  
 Db 1953 ACCTGTCTTGAATCTCTGGCGGAGAGAGAGCTGACAGCGGTGAGTCTTGCTA 2012  
 QY 3536 CGCAGCGCGGCTTGTGAGCTCGGCAAGCGGACTGTTACGGGATACCAAGCTCGGCG 3595  
 Db 2013 CGCAGCGGCTCTCTGGAATTTGGCAATTTCGACTTTCAGAACACCGCGCTCGGCA 2072  
 QY 3596 TGGCGCGGCTCTCGGCAATCTCTCTCTGCTGTGGATCTCCGGGGATGATGCTCG 3655  
 Db 2073 TGGCTATCTTCTTAAGAGAGAGTTCACATTCACGCGGCTTCACTGATGCTTTCACAG 2132  
 QY 3656 AGCGCGCGCGCGGCTCGGCTCTTGAAGAGAGCTCTCGGCTGATCGCGGAGAGAG 3715  
 Db 2133 AGACAGTGTGACTGTGCGGAGAGGTGTGGCGCTTGTGACAGCGCGGATCCGGATGCGG 2192  
 QY 3716 TGTTCACCCCTCCCGCCATCGGACGCTCCGATCGCGGCTGTGCGCGATGCTTCGGA 3775  
 Db 2193 TGTACGCGCCCTCAAGTGCACGAGTGTTCATGAGGCGCCAGGTGAGAGACGCTTCGCT 2252  
 QY 3776 GCATGGCGCAGCGCCAG 3792  
 Db 2253 ACATGGCCCAAGGAG 2260

## RESULT 11

B0069153 1061 bp mRNA linear EST 02-APR-2002  
 LOCUS B0069153 AGENCOURT 6738542 NIH\_MGC\_47 Homo sapiens CDNA clone IMAGE:5803443  
 DEFINITION 5', mRNA sequence.  
 ACCESSION B0069153  
 VERSION B0069153.1 GI:19898199  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1061)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.







Db 1379 GCCCGGGGNCGCCGNCNCGCGCGCGCGCCGCGCCNCCGCGNCGCGCGGCGGNC 1438  
 Oy CGAGTCGATGAGCCAGCGTGTTCATCCTCGTGTTCGGGTCACGAGCGCGCGC 2909  
 Db 1439 CGGCGCGCNCGCNC 1498  
 Oy 2910 CCTGTGTGGGCGAGGTCGAGATCCGCGTGA CGCGCGCGGAGCTCAGCTTCAATGATG 2969  
 Db 1499 CGGCGCGCGCNCNCCGNNCCGNNCCGCGCGCGCNCNCGCGCNCNCCGCGCG 1558  
 Oy 2970 CCACTCGCGCTGGGCGATGTGTCCCGATGACCTGCGGGAAAGCCCAACCTCGCTCT 3029  
 Db 1559 CGGCN 1615  
 Oy 3030 GCTGGAGGCGAGTGTGCGCGCGCGCGCATGTGCG -CGTGGGCGAGGAGCGTGAACGCGCTCG 3088  
 Db 1616 CGCGCGCNCNCG 1675  
 Oy 3089 TGTGTGGGCGAACCGGTCAATCGCCCTTTTCGGCGGAGCGCTTTGTCAACCACTCAACGT 3148  
 Db 1676 CGCGCGCNCG 1735  
 Oy 3149 CGGCTGCGCTGTGTGCTGCTCGGCTCAAGCGCTCTCGGCGATGAGGCGCGCGCATGC 3208  
 Db 1736 GCNC 1795  
 Oy 3209 CC 3210  
 Db 1796 GC 1797

RESULT	14
BZ560333	
LOCUS	BZ560333
DEFINITION	BZ560333 599 bp DNA linear GSS 17-DEC-2002 pac82-164_2364.x1 pac82-164 Pseudomonas aeruginosa genomic clone pac82-164_2364, genomic survey sequence.

VERSION	BZ560333.1
KEYWORDS	GI:21718168
SOURCE	GSS.
ORGANISM	<i>Pseudomonas aeruginosa</i> <i>Pseudomonas aeruginosa</i> <i>Bacteria</i> ; <i>Proteobacteria</i> ; <i>Gammaproteobacteria</i> ; <i>Pseudomonadales</i> ; <i>Pseudomonadaceae</i> ; <i>Pseudomonas</i> .
REFERENCE	1 (bases 1 to 599)
AUTHORS	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE	Whole-Genome-Sequence Variation among multiple isolates of <i>Pseudomonas aeruginosa</i> library
JOURNAL	<i>J. Bacteriol.</i> , (2002) In press
COMMENT	Contact: Chris K. Raymond

University of Washington  
Box 353145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: [craymond@u.washington.edu](mailto:craymond@u.washington.edu)  
Class: shotgun.

```

FEATURES
  source
    Location/Qualifiers
      1..599
        organism="Pseudomonas aeruginosa"
        mol_type="genomic DNA"
        strain="2-164"
        db_xref="taxon:287"
        clone="pacs2-164_2364"
        clone_1fb="pacs2-164"
        note="Clinical isolate 2-164 Whole genomic shotgun
library."

```

	BASE COUNT	ORIGIN	89 a	184 c	199 g	126 t	1 others
Query Match	2.7%						
Best Local Similarity	52.2%						
Pred.No.	2.2e-10;						
Score	114.4;						
DB	29;						
Length	599;						

Matches	253;	Conservative	0;	Mismatches	232;	Indels	0;	Gaps	0;
QY	34	GATCGATTGCGATCGTCGGAGCGAGTTCGCTTCGCGCGGTGGCGGTGATCGATCTTGAGC	93						
Db	111	GAACCCCGTCGAATCATCGGAGCGGCGCTTGCCTGTTTCCGGCGGAGGTGCTGCGCCGGAG	170						
QY	94	GGGTTCTTGACGCTCTCTCGAAGGCTCGCGGACACCGTGGGGCGAGTCTCCCGCGAAAGC	153						
Db	171	GCATTCTTGAAACCTCTTCGTGTCGTGGACCGAATTTGTCGGGCGCGTCCCGCGGCGCG	230						
QY	154	TGGGATGACGAGCGTGGTTGATCCGACCCCGATGCTCCCGGGAGAGACCGCCGTAAAC	213						
Db	231	TGGATGCGCGAGGTCGATTCAACGCGCTGTGTGCGAGCCGTCCTCCGGTACCATCTATTGCCGG	290						
QY	214	CGCGCATCTTCTCTGACCGCAGTGAAGCTGTCTCGACGCTCTCTTTCGCGATCTCGCT	273						
Db	291	GAGGGGGCGCTTCCTGTATGAGTGGAGCCGCTGACGCGGGTTCCTTCGGGATATCCGCT	350						
QY	274	CGCGAAGCGCTCGCGATGACCTTCGACATCGACTCTTGTGGAAGTGTGCTGGAGAGCG	333						
Db	351	TCCGAGGCTCGCGAAATGGAACCCCGACACCGCTGTGCTGTGAGGTCGTCTGGGAATGC	410						
QY	334	CTGGAGAACGCGCGATCGCTCCATCGGCGGCTCGTCGTGCGGAAACGGGAGTGTTCATC	393						
Db	411	ATGGAACGGGCGGGGCTTGTCTCTGAAACGCTGAAAGTTTCCCGCATCGGGCTTACGTC	470						
QY	394	GGGATCGGCGCGTCGGAATATGAGGCCGCGCTGCGCGAACGACGGCGTCCGACAGATC	453						
Db	471	GGCATGTGCGGCATGAGTTACCTTACCCTGCGATGCGCGAGGCGGGTATCCCGCGGATC	530						
QY	454	GACGCTATGCGGGGCTGGGGACGATCCCAAGCTCGGACCGGGCCGAATCTGTATGCC	513						
Db	531	AATCATCTTATATGCGGCTGGANGGAGTTGAGCTTCGCGCCGCGCGCATCTGCTATCAC	590						
QY	514	CTGGG 518							
Db	591	CTTGG 595							

RESULT 15  
 BUS28700  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

BUS28700 899 bp mRNA linear EST 13-SEP-2002  
 AGENCOURT 10183164 NIH MGC 101 Homo sapiens cDNA clone  
 IMAGE:6538675 5', mRNA\_sequence.  
 BUS28700  
 BUS28700.1 .GT:22839141  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 899)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [crabbe-remail.nih.gov](mailto:crabbe-remail.nih.gov)

Tissue: Urogenital  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Robin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MCC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2703 row: j column: 19  
 High quality sequence stop: 600.  
 Location/Qualifiers  
 1. 899  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6538675"  
 /tissue\_type="epidermoid carcinoma, cell line"

```

/lab.host="DH10B (phage-resistant)"
/clone.lib="N1H1 MGC 101"
/notes="Organ: lung; Vector: pGB7; Site_1: EcoRI; Site_2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAGC(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
N1H1 MGC library."

```

BASE COUNT	145 a	280 c	307 g	166 t	1 others
ORIGIN					

Query Match	2.7%;	Score 113.4;	DB 13;	Length 899;
Best Local Similarity	49.4%;	Pred. No. 3.3e-10;		
Matches 358;	Conservative 0;	Mismatches 352;	Indels 14;	Gaps 2;

OY	3128	TTGCTAACCCAGTACACAGCTCGAGTGGAGCTGATCTCGCTCGAGCGAGGCTCTGG	3187
Db	46	TGGCCACTCTGTCTGCTGTACCGGACTTCTCTTGGGAATGTGCTTTCACACTGGAACGC	105
OY	3188	CGATTCGAGCGGAGCGGCATTCGCGCGTACCTGACGGCATGTGATCGCGCTCGACAGA	3247
Db	106	TGAGAGAGCGGCGCTCGGTGCTGTGCTGCTACAGACGGGCTACTACGGCGGTGTGTC	165
OY	3248	TAGCCCGCTTTCAGCCGGGGGAGCGGGTGTGATTCATGCGCGCACCGCGGGGTGGTC	3307
Db	166	GTGGGCGGGGTACCGCCCGGGGAGACGTGTCTCATCTACCTCGGAGCTCGGGCGGCTGGAGCC	225
OY	3308	TGCGCGCGGTGCAATGAGGAGCGGACGACGATGGGAGCGGAGTGCATATGACGCGCGGACAGC	3366
Db	226	AGGCGCGCATGCCATCGCCCTCACTGTGGGTGTGCGGCTTTTACCAACCTGGGTGG	285
OY	3368	CCGAGAAACGCGCTTACCTGAGTGTGCTGGGCGTGGCTGATGTGAG-----CG	3415
Db	286	CTGAGAAACGGGCGTACCTCCAGGCGCAGATTTCCCACTGTGACAGCACAGCTTCGGCA	345
OY	3416	ATTCCCGCTCGGACCGGTTTCGTGCGCCACGTGCGCGCTGACGCGGCGCGAGGAGTGG	3475
Db	346	ACTCCCGGGACACATCTTTCGAGCAGCATGTGCTGTGCAACCGGCGGGAAAGGCGTTG	405
OY	3476	ACGTGCTCTCAACTCGCTCTCGGAGCAGCTGATGACAGAGTTTCAATCTCCGTGCAT	3535
Db	406	ACCTGTCTTGAATCTCTTGGCGGAGAGAAAGCTGACAGCCAGCTGAGTCTTGCTTA	465
OY	3536	CGCAGCGCGGTTGTGTGAGCTCGGCAAGCGCGACTGTTAACGCGGATAACAGACTCGGC	3595
Db	466	CGCAGGTGTGCTCTGTGAATTTGGCAATTTGACTTTTTCAGAACCAACCGCTCGGCA	525
OY	3596	TGCGGCGTTCCTGTGCAATCTCTCTTCTGTGCTGTGATCTCCGGGGATGATGCTCG	3655
Db	526	TGGCTATCTTCTTGAAAGACGTGACATTTCAACGGGGTCTTACTGATCGTTTTCACAG	585
OY	3656	AGCGCGCGGCGCGGTCGCTCGGTGCTCTTGGAGAGACTCTCGGCTGATTCGCGGAGCGC	3715
Db	586	AGAGCAGTGTGACTGTGCGGGAGAGGTGTGGCGCTTGTCAAGCGCGGCATTCGGGATGGGG	645
OY	3716	TGTTTCAACCCCTTCCCCCATTCGCGAGCTGCC--GATTCGCGCGTGTGCGCGATGCGTTCCG	3773
Db	646	TGTGATCGGGCCCTCAAGTGCACAGGTGTTCCATGGGGGCCAAGTGGAGAGACGCTTCCG	705
OY	3774	GAGCATGGCGAGGCGCAGCATCTTTGGAGACTGTACTCAAGCTGGGTGAACCGGAGGT	3833
Db	706	CTACTGTGCCAGGGGAAACACATTTGGCAAAAGTGCCTGACAGGGGCTTGCCTGGAGAA	765
OY	3834	CCAG 3837	
Db	766	CCCG 769	

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using BW model

Run on: October 4, 2003, 03:31:13 ; Search time 981.368 Seconds  
(without alignments)  
11088.586 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_7610\_11875

Perfect score: 4266  
Sequence: 1 gtggcgagatcgcccatcga.....gggaatcatcgcccatga 4266

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4266	100.0	68750	13	US-10-014-717-1
2	814.2	19.1	3978	12	US-09-953-348-19
3	814.2	19.1	3978	14	US-10-267-255-19
4	814.2	19.1	12249	12	US-09-953-348-74
5	814.2	19.1	12249	14	US-10-267-255-74
6	814.2	19.1	18331	12	US-09-953-348-96
7	814.2	19.1	18331	14	US-10-267-255-96
8	770.8	18.1	5355	14	US-10-205-032-19
9	770.8	18.1	60196	14	US-10-205-032-1
10	765.8	18.0	6360	14	US-10-212-962-2
11	761.2	17.8	10839	14	US-10-156-761-2882
12	761.2	17.8	125746	14	US-10-156-761-15102
13	761.2	17.8	9025608	14	US-10-156-761-1
14	742	17.4	15872	10	US-09-861-289-1
15	742	17.4	15872	10	US-09-860-846-1
16	742	17.4	15872	11	US-09-988-384B-1

17	742	17.4	15872	11	US-09-836-821-1	Sequence 1, Appl
18	732.8	17.2	18435	14	US-10-156-761-412	Sequence 412, App
19	732.8	17.2	100000	14	US-10-156-761-15103	Sequence 15103, A
20	715.8	16.8	18438	14	US-10-156-761-2886	Sequence 2886, Ap
21	707	16.6	23238	14	US-10-156-761-415	Sequence 415, App
22	695.4	16.3	11817	14	US-10-156-761-2884	Sequence 2884, App
23	679.6	15.9	10692	14	US-10-156-761-414	Sequence 414, App
24	679.4	15.7	14427	14	US-10-205-032-15	Sequence 15, Appl
25	669.2	15.6	15427	14	US-10-156-761-1540	Sequence 1540, App
26	667.6	15.6	5505	14	US-10-156-761-413	Sequence 413, App
27	666	15.6	14520	14	US-10-156-761-2885	Sequence 2885, Ap
28	655.6	15.4	10056	14	US-10-156-761-411	Sequence 411, App
29	652.6	15.3	11910	14	US-10-156-761-2879	Sequence 2879, App
30	652.6	15.3	125746	14	US-10-156-761-15102	Sequence 15102, A
31	652.6	15.3	9025608	14	US-10-156-761-1	Sequence 1, Appl
32	648.6	15.2	4725	14	US-10-205-032-17	Sequence 17, Appl
33	633	14.8	77536	12	US-09-940-3168-1	Sequence 1, Appl
34	627.8	14.7	13494	14	US-10-156-761-2355	Sequence 2355, Ap
35	627.6	14.7	13842	10	US-09-861-289-30	Sequence 30, Appl
36	627.6	14.7	13842	10	US-09-860-846-30	Sequence 30, Appl
37	627.6	14.7	13842	11	US-09-988-384B-30	Sequence 30, Appl
38	627.6	14.7	13842	11	US-09-836-821-30	Sequence 30, Appl
39	627.6	14.7	36778	10	US-09-861-289-5	Sequence 5, Appl
40	627.6	14.7	36778	10	US-09-860-846-5	Sequence 5, Appl
41	627.6	14.7	36778	11	US-09-836-821-5	Sequence 5, Appl
42	627.6	14.7	37948	11	US-09-988-384B-19	Sequence 19, Appl
43	627.6	14.7	38506	11	US-09-793-708-19	Sequence 19, Appl
44	627.6	14.7	38506	12	US-10-201-365-1	Sequence 1, Appl
45	627.6	14.7	38506	12	US-10-160-539-19	Sequence 19, Appl

## ALIGNMENTS

RESULT 1  
US-10-014-717-1  
; Sequence 1, Application US/10014717  
; Publication No. US20020192778A1  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Liqon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/10/014, 717  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US/09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 68750  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-10-014-717-1  
Query Match 100.0%; Score 4266; DB 13; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGGCGATCTCTCCATCGAGCGCGGACCGAAGATCGATTGGATGTCGGAGCGAGT 60  
DB 7610 GTGGCGATCTCTCCATCGAGCGCGGACCGAAGATCGATTGGATGTCGGAGCGAGT 7669  
QY 61 TGCGGTGCGCGGATGCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 120  
DB 7670 TGCGGTGCGCGGATGCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 7729  
QY 121 CGGACACGTCGGCGGAGTCGCCCGGACGCTGGAGTCGACGACGAGCGTGTGATCCC 180

Db 7730 GCGGACACGCTGAGGAGTCCCCCGGACAGCGCTGGATGACGACGCTGTTATCCC 7789  
Qy 181 GACCCGATGCCCGGGGAGAGCGCCGTTACGCGCGCATCTTTCTTGAGCAGCTAGCC 240  
Db 7790 GACCCGATGCCCGGGGAGAGCGCCGTTACGCGCGCATCTTTCTTGAGCAGCTAGCC 7849  
Qy 241 TGCCTGACGCGCTCTTCTTGAGCATCTGCGGCGGAGGCGCTGGATGAGCCCTGCA 300  
Db 7850 TGCCTGACGCGCTCTTCTTGAGCATCTGCGGCGGAGGCGCTGGATGAGCCCTGCA 7909  
Qy 301 CATGCACTCTTGCTGAGAGTGTGCTGAGAGCGCTGAGAGACGCGCGATCGCTCATCG 360  
Db 7910 CATGCACTCTTGCTGAGAGTGTGCTGAGAGCGCTGAGAGACGCGCGATCGCTCATCG 7969  
Qy 361 GCGCTGCTGCTGAGAGGAGGAGTGTTCATCGGAGATCGGCCGCTCGAATATGAGGCC 420  
Db 7970 GCGCTGCTGCTGAGAGGAGGAGTGTTCATCGGAGATCGGCCGCTCGAATATGAGGCC 8029  
Qy 421 GCGCTGCGGACGACGCGCGCTCGGAGAGTGAAGCTCATGAGGCGCTGGGAGCGATG 480  
Db 8030 GCGCTGCGGACGACGCGCGCTCGGAGAGTGAAGCTCATGAGGCGCTGGGAGCGATG 8089  
Qy 481 CCGACGCTGAGAGCGGCGCGAATCTGTATGCCCTCGGCGCTGAGAGCGCGCTGTGCGG 540  
Db 8090 CCGACGCTGAGAGCGGCGCGAATCTGTATGCCCTCGGCGCTGAGAGCGCGCTGTGCGG 8149  
Qy 541 GTGATAGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Db 8150 GTGATAGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8209  
Qy 601 TCCCGGAGATGCTCCACGCGCGCTGCGTGGTGGGATGATGCTGATGTTGCGGAGCAC 660  
Db 8210 TCCCGGAGATGCTCCACGCGCGCTGCGTGGTGGGATGATGCTGATGTTGCGGAGCAC 8269  
Qy 661 CTGCTGCTGCTGCTGCAAGACCGCGCGCTGCGCAGGAGCGGTGCTGCAAGGACATTTTCG 720  
Db 8270 CTGCTGCTGCTGCTGCAAGACCGCGCGCTGCGCAGGAGCGGTGCTGCAAGGACATTTTCG 8329  
Qy 721 GCGGAGGCGGATGAGGTTGAGACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
Db 8330 GCGGAGGCGGATGAGGTTGAGACGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8389  
Qy 781 AGTGAAGCGCGCGCGGAGCGGCGATGAGATATGAGGAGTGAATTCGAGGATCCGCGATCA 840  
Db 8390 AGTGAAGCGCGCGCGGAGCGGCGATGAGATATGAGGAGTGAATTCGAGGATCCGCGATCA 8449  
Qy 841 CACGACGCTGCGAGACGCGGCTGTAACGCTGCGGAAACGAGAGCTCCCAAGAAATCGTCTG 900  
Db 8450 CACGACGCTGCGAGACGCGGCTGTAACGCTGCGGAAACGAGAGCTCCCAAGAAATCGTCTG 8509  
Qy 901 AAAACGCGGCTGCGGAGCGGAGGCTGCGCGGCTGCTGCGTGGGTTATGTCAGAGCACAC 960  
Db 8510 AAAACGCGGCTGCGGAGCGGAGGCTGCGCGGCTGCTGCGTGGGTTATGTCAGAGCACAC 8569  
Qy 961 GCGACGCGGACGACGCTTGTGTAACCCATCGAAATTCGAAGCTCTGATATGCGATATGCGC 1020  
Db 8570 GCGACGCGGACGACGCTTGTGTAACCCATCGAAATTCGAAGCTCTGATATGCGATATGCGC 8629  
Qy 1021 CTGCGGCGGAGATGTCGCAACGCGCTGCTGATCGGAGTGGTGAAGACCAACCTTGCCAT 1080  
Db 8630 CTGCGGCGGAGATGTCGCAACGCGCTGCTGATCGGAGTGGTGAAGACCAACCTTGCCAT 8689  
Qy 1081 CTTGATGATGCTGCGGAGTCACTGAGGCTGCTGAAGGCTGCTTGTCCCTTACGACGCGG 1140  
Db 8690 CTTGATGATGCTGCGGAGTCACTGAGGCTGCTGAAGGCTGCTTGTCCCTTACGACGCGG 8749  
Qy 1141 CAGATTCCTGCGACCTCCACGCGGAGGCGCTGAAGCCCGGATCTCATGAGGAGTATCTT 1200  
Db 8750 CAGATTCCTGCGACCTCCACGCGGAGGCGCTGAAGCCCGGATCTCATGAGGAGTATCTT 8809  
Qy 1201 CCGCTGACCGCTGACGCGCGCGCGGACACCGTGGCGGAGCTGGAATAGCGCGGAGCGGCG 1260  
Db 8810 CCGCTGACCGCTGACGCGCGCGCGGAGACCGTGGCGGAGCTGGAATAGCGCGGAGCGGCG 8869

Qy 1261 GGGGTGACCTGCTTCCGATGAGCGGAGCAACCGCGACGCTGCTGGAAGAGGCGCG 1320  
Db 8870 GGGGTGACCTGCTTCCGATGAGCGGAGCAACCGCGACGCTGCTGGAAGAGGCGCG 8929  
Qy 1321 GCGGAGGAGTCAACCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380  
Db 8930 GCGGAGGAGTCAACCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8989  
Qy 1381 AGGACCGGCTACGCCCTGAGATGACAGGCGCGCGCGCTGCGGACCATCTGAGACCTTAC 1440  
Db 8990 AGGACCGGCTACGCCCTGAGATGACAGGCGCGCGCGCTGCGGACCATCTGAGACCTTAC 9049  
Qy 1441 CTTTCCGAGTGTCTGAGGAGATGAGGCGCTTCACTGCGGAGAGGAGGAGGAGGAGGAG 1500  
Db 9050 CTTTCCGAGTGTCTGAGGAGATGAGGCGCTTCACTGCGGAGAGGAGGAGGAGGAGGAG 9109  
Qy 1501 CACCGGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560  
Db 9110 CACCGGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9169  
Qy 1561 GCGGAGGAGACAGCGTCCCGGCTGCGGAGTGTGAGTATGCTGATCTTCAACGCGGCAAG 1620  
Db 9170 GCGGAGGAGACAGCGTCCCGGCTGCGGAGTGTGAGTATGCTGATCTTCAACGCGGCAAG 9229  
Qy 1621 CTGCGCTTCTTCTTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680  
Db 9230 CTGCGCTTCTTCTTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9289  
Qy 1681 GATGATGATGCTCGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740  
Db 9290 GATGATGATGCTCGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9349  
Qy 1741 CTGACCGGCGGCTTCCGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800  
Db 9350 CTGACCGGCGGCTTCCGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9409  
Qy 1801 CTGACCGGAGACGCTTACCCAGCGCGCGCTGTTTCACTTCTGAAATAGCGCTGCGCGG 1860  
Db 9410 CTGACCGGAGACGCTTACCCAGCGCGCGCTGTTTCACTTCTGAAATAGCGCTGCGCGG 9469  
Qy 1861 CTGAGCGGCTGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920  
Db 9470 CTGAGCGGCTGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9529  
Qy 1921 GTGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980  
Db 9530 GTGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9589  
Qy 1981 CCGCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040  
Db 9590 CCGCGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9649  
Qy 2041 GAGGCGGATGAGGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100  
Db 9650 GAGGCGGATGAGGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9709  
Qy 2101 AACGCTCCGAGACAGAGTGTATGAGCGGCGCGCGGAGCAACCGTGTGATGCGAGGCGG 2160  
Db 9710 AACGCTCCGAGACAGAGTGTATGAGCGGCGCGCGGAGCAACCGTGTGATGCGAGGCGG 9769  
Qy 2161 GCGATGAGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220  
Db 9770 GCGATGAGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9829  
Qy 2221 CCGCTCATGAGGCGCGGATGCTGAGAGGCTTCCGAGGAGTGTGAGGAGTGTGAGGAGG 2280  
Db 9830 CCGCTCATGAGGCGCGGATGCTGAGAGGCTTCCGAGGAGTGTGAGGAGTGTGAGGAGG 9889  
Qy 2281 CCGCGCTGATGCTGCTGTCAGCAATTTAGCGGAGGAGGCTTGTGACAGAGAGGTGAGC 2340  
Db 9890 CCGCGCTGATGCTGCTGTCAGCAATTTAGCGGAGGAGGCTTGTGACAGAGAGGTGAGC 9949

Qy	2341	TCGCGCGGCTATTGGGTCGCGCACGCGCGAGAGGTGGCTCTCGCGATGTGAAG	2400
Db	9950	TCGCGCGGCTATTGGGTCGCGCACGCGCGAGAGGTGGCTCTCGCGATGTGAAG	10006
Qy	2401	GCGCTGCACGCGCGCGGTGCGGACCTTCGTGAGGTGGTCCGAATGACGCTGTC	2460
Db	10010	GCGCTGCACGCGCGCGGTGCGGACCTTCGTGAGGTGGTCCGAATGACGCTGTC	10066
Qy	2461	GCGCTGGTGGCTGCTGCTTCGATGTCGGAGACGCCCGGCGTGGCTGTGTGGCGCT	2520
Db	10070	GCGCTGGTGGCTGCTGCTTCGATGTCGGAGACGCCCGGCGTGGCTGTGTGGCGCGCT	10126
Qy	2521	GAGGCTGACGAGCGCGGCGACCGTGTCTGAGAGCGCTCGCGGAGCTGAGGCGGTGGCG	2580
Db	10130	GAGGCTGACGAGCGCGGCGACCGTGTCTGAGAGCGCTCGCGGAGCTGAGGCGGTGGCG	10186
Qy	2581	CTGGTCTCTTGGGCGGCGCTCTTCCCTTCAGGGGGGGCGGCGGGTGCCTGCCACGTAC	2640
Db	10190	CTGGTCTCTTGGGCGGCGCTCTTCCCTTCAGGGGGGGCGGCGGGTGCCTGCCACGTAC	10246
Qy	2641	CTTTGGCAGCGCGGACCGCTTACTGATGACACGAAAGCCGACGACGCGCGCTGGCGAC	2700
Db	10250	CTTTGGCAGCGCGGACCGCTTACTGATGACACGAAAGCCGACGACGCGCGCTGGCGAC	10306
Qy	2701	CGCGGTCTCCGGGAGGGGGTACAGAGAGTGGAGAGGGGGGGCGGCGGCGCGCGCG	2760
Db	10310	CGCGGTCTCCGGGAGGGGGTACAGAGAGTGGAGAGGGGGGGCGGCGGCGCGCGCGCG	10366
Qy	2761	GACCGGCGCAGCGCTCGGCTGCACATCCGCGCGCGCGAGACGGAACGCGCGGAGAAAGTC	2820
Db	10370	GACCGGCGCAGCGCTCGGCTGCACATCCGCGCGCGCGAGACGGAACCGCGGAGAAAGTC	10422
Qy	2821	GAGGCGCGCGGCGACCGTCCGTTCCGAGCTCGAGATCGATGAGCCAGCGCTGTGATCAC	2880
Db	10430	GAGGCGCGCGGCGACCGTCCGTTCCGAGCTCGAGATCGATGAGCCAGCGCTGTGATCAC	10488
Qy	2881	CTCGTGGTTGGGGTCAAGAGACGGCGGGCGCCCTGGTGTGGGCGAGGTCCAGATCGCGCTTC	2940
Db	10490	CTCGTGGTTGGGGTCAAGAGACGGCGGGCGCCCTGGTGTGGGCGAGGTCCAGATCGCGCTTC	10546
Qy	2941	GACCGGCGGCGGCTCAGACCTTCAATGATGTCAAGTCGCGCTGGGACATGATGCCGACGAC	3000
Db	10550	GACCGGCGGCGGCGCTCAGACCTTCAATGATGTCAAGTCGCGCTGGGACATGATGCCGACGAC	10606
Qy	3001	CTGCGGAGAAAGCCCAACCTTCGCGTGTCTCGAGAGCGATGCGCGGCGCGCATGCTC	3060
Db	10610	CTGCGGAGAAAGCCCAACCTTCGCGTGTCTCGAGAGCGATGCGCGGCGCGCATGCTC	10666
Qy	3061	GCCGTGGGCGGAGGCGGTGAACCGCTCTGTGTGGGCGCAACCGGTCACTCGCCCTTTGGCG	3120
Db	10670	GCCGTGGGCGGAGGCGGTGAACCGCTCTGTGTGGGCGCAACCGGTCACTCGCCCTTTGGCG	10726
Qy	3121	GGAAGCTTTGTGACCCACGTCACCAACGTCGAGTCGCGTGGTGCGCTGGGCGCTCAGGGCG	3180
Db	10730	GGAAGCTTTGTGACCCACGTCACCAACGTCGAGTCGCGTGGTGCGCTGGGCGCTCAGGGCG	10786
Qy	3181	CTCTCGGCGATTCGAGAGCGGCGCGCATGCGCGCTGCATTCACGCGATGATACGCGCTC	3240
Db	10790	CTCTCGGCGATTCGAGAGCGGCGCGCATGCGCGCTGCATTCACGCGATGATACGCGCTC	10846
Qy	3241	GACGAAATAGCCCGCTTTCAGCGCGGGGAGCGGGTGTGATTCATGTGGCGGACCGCGCGG	3300
Db	10850	GACGAAATAGCCCGCTTTCAGCGCGGGGAGCGGGTGTGATTCATGTGGCGGACCGCGCGG	10906
Qy	3301	GTCGAGTCTCGCGCGGCGGTGACAGTGGGCGCAGACGTCGGAGCCGAGGTCCATGTGAGAGGCG	3360
Db	10910	GTCGAGTCTCGCGCGGCGGTGACAGTGGGCGCAGACGTCGGAGCCGAGGTCCATGTGAGAGGCG	10966
Qy	3361	GGCACGCGCGGAGAAACGCGCTTACCTTGAAGTCGCTGGGCGTGGCGTATGTGACGATTC	3420
Db	10970	GGCACGCGCGGAGAAACGCGCTTACCTTGAAGTCGCTGGGCGTGGCGTATGTGACGATTC	11022
Qy	3421	CGCTCGGACCGGTTGTCGCCGACGTGCGCGCTGCACGCGGCGCGAGGAGTAAAGCTC	3480

Db	11030	CGCTCGGACCGGATTTCGTCCGCCGACGTCGCGCGCGTGGACGGCGGGGAGAGAGTAGACGTC	11089
OY	3481	GTGCTCAACTCGCTCTCGGGCGCACTGATCGACAAAGATTTCATCTCTCGCATCGCAC	3540
Db	11090	GTGCTCAACTCGCTCTCGGGCGCACTGATCGCAAGATTTCATCTCTCGCATCGCAC	11149
OY	3541	GGCCGGTTGTGGAGCTCGGCAAAAGGCGCATGTCTTACGGGAAATACCACTGGGCTGGCG	3600
Db	11150	GGCCGGTTGTGGAGCTCGGCAAAAGGCGCATGTCTTACGGGAAATACCACTGGGCTGGCG	11209
OY	3601	CCGTTCTGCGCAATCTCTTCTTCGCTGTGTGATCTCCGGGGATGATGCTCGAGCGG	3660
Db	11210	CCGTTCTGCGCAATCTCTTCTTCGCTGTGTGATCTCCGGGGATGATGCTCGAGCGG	11269
OY	3661	CCGACCGGGGTCCGTGCGCTCTTTGAGAGACTCTCGGCTGATGCGGCAAGCGTGTTC	3720
Db	11270	CCGACCGGGGTCCGTGCGCTCTTTGAGAGAGCTCTCGGCTGATGCGGCAAGCGTGTTC	11329
OY	3721	ACCCCTCCCCCAATGCGGACGCTCCCGAATCGGCGGCTGCGGATGCGGTCGAGACATG	3780
Db	11330	ACCCCTCCCCCAATGCGGACGCTCCCGAATCGGCGGCTGCGGATGCGGTCGAGACATG	11389
OY	3781	GCGCAGGCGCAGCATCTTGGGAAAGCTCGTACTCAGCTGGGTGACCCGAGGTCCAGATC	3840
Db	11390	GCGCAGGCGCAGCATCTTGGGAAAGCTCGTACTCAGCTGGGTGACCCGAGGTCCAGATC	11449
OY	3841	CGTATTCCAAACCAACGACGACGCGCGGCTCGTACACCGGGGATCGGGAACCTGTCGACAGG	3900
Db	11450	CGTATTCCAAACCAACGACGACGCGCGGCTCGTACACCGGGGATCGGGAACCTGTCGACAGG	11509
OY	3901	CTGGCGTCAGCTGCGCGCGGCGCGCGGCGGGGCGGTGGAGGCGTTCTCGTACGCGAG	3960
Db	11510	CTGGCGTCAGCTGCGCGCGGCGCGCGGCGGGGCGGTGGAGGCGTTCTCGTACGCGAG	11569
OY	3961	GTCCTCAGAGTGTGCGCACGCGCCGCAATCAAGTCTGGGCGCGAGCGCTGTTCACCCGC	4020
Db	11570	GTCCTCAGAGTGTGCGCACGCGCCGCAATCAAGTCTGGGCGCGAGCGCTGTTCACCCGC	11629
OY	4021	CTGGCGCATGGAATCTGCTCATGCGCCGTGGAGCTGCGCAATCGTATGAGGCGAGCTCAAG	4080
Db	11630	CTGGCGCATGGAATCTGCTCATGCGCCGTGGAGCTGCGCAATCGTATGAGGCGAGCTCAAG	11689
OY	4081	CTGAAGCTGTCAGACGAGTTCCTGTCACAGTCCGCCAATATGAGCTTGTGGAGCCAAAC	4140
Db	11690	CTGAAGCTGTCAGACGAGTTCCTGTCACAGTCCGCCAATATGAGCTTGTGGAGCCAAAC	11749
OY	4141	CTGTGGATGCTCTTCGCAACAGCTCTTCTTTGAGCGGAGTGGCGGCGAGAACCTACGG	4200
Db	11750	CTGTGGATGCTCTTCGCAACAGCTCTTCTTTGAGCGGAGTGGCGGCGAGAACCTACGG	11809
OY	4201	GCGAGGCTGCAAAACGACTTTCGTCTCATGCGGCGCAGATCAAGACTGGGAAATCATTTGCC	4260
Db	11810	GCGAGGCTGCAAAACGACTTTCGTCTCATGCGGCGCAGATCAAGACTGGGAAATCATTTGCC	11869
OY	4261	CTATGA 4266	
Db	11870	CTATGA 11875	

```

US-RESULT 2
US-09-953-348-19
; Sequence 19, Application US/09953348
; Publication No. US2003013498A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqiang
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600 530U51
; CURRENT APPLICATION NUMBER: US/09/953.348

```

```

:
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: PCT/US00/06394
:
: PRIOR FILING DATE: 2000-03-10
: PRIOR APPLICATION NUMBER: 09/266965
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 153
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 19
:
: LENGTH: 3978
:
: TYPE: DNA
:
: ORGANISM: Streptomyces lavendulae
:
: US-09-953348-19

```

Query Match	19.1%	Score 814.2;	DB 12;	Length 3978;
Best Local Similarly	57.6%	Pred. No. 7.7e-170;		
Matches 1549; Conservative	0;	Mismatches 1108;	Indels 30;	Gaps 4;

Qy	18	CGAGCGGACGCGCGAAGATTCGATTCGATTCGTCGAGGAGAGTTTCGCTCGCGCGGTTG	77
Dp	90	CGAGAGACGCGGCCCCGCGAGCCGTCGCATTCGTGGGCATAGCGTGCCTCTCCCGGGGA	149
Qy	78	CGTATCGATCTGAGCGGGTCTGACGCTCTCTGAGGAGCTGCGCGACACCGTGGAGC	137
Dp	150	CGTGGCATCGCGGAGCGACTGTGGCAGCTGTGTCCGAGGGGCGGAGCGCGTCAACGA	209
Qy	138	AGTCCCGCGCGAAC---GCTGGGATGCAACACGTGGTTTATCCCGACCCCGATGCCCC	194
Dp	210	GTTCCCGCGCGACCGGGGCTGGGACGTCAACGCGCTTACAGACCCCGAGCGGGCACCCC	269
Qy	195	GGGGAAGACGCGCGTTACGCGCGCATCTTCTTGACGACGATGACTTGTCGACGCTC	254
Dp	270	GGGCAAGACGTACGCGCGCGCACGCGCGCTTCTTCAAGAGACGCGCGGATTCGACGCGC	329
Qy	255	CTTCTTGGCATCTTCGCTTCGCGAAGCGCTGCGATGACCTCTGACATGATCTTTGCT	314
Dp	330	CTTCTTGGCATCACCCCGCGAGCGCTTCGCATGACCCGACGAGCGCATGATCAT	389
Qy	315	GGAGGTGTGTGGGAGCGCTGAGAGACGCGCATGCTCATTCGGCGCTGTGCTATC	374
Dp	390	GGAGGTCTCTGGGAGGCGTTGAGAGGCGGGGCTTCACGCGACACCTTCGCGGGCGA	449
Qy	375	GGAAACGGAGTGTTCATCGGATCGGCGCGCTCCGATATGAGGCGCGCTCGCGCAAGC	434
Dp	450	GGACGTGGGTCTTGTGTGGCTTCCAAACGACAGCATACGTATCAACGTGTCTGACGC	509
Qy	435	GACGGCGTCCGACAGATGAGGCTCATGCGCGGCTGGGACGATGCCAGCGTGGAGC	494
Dp	510	GCGGAGGTGCGCGAGGGGCTTC-----ATCGGAGCGGCAACTCCGCAACATCTCTC	563
Qy	495	GGGCGCAATCGTATGACCTCGGGCTGCGAGGGCGGTGTGCGCGGTGATACGGACTA	554
Dp	564	CGGCGCGGTGCTTACACTTTCGACTTCGAGGAGCCGGGCGTGTCTGTGACACCGCTG	623
Qy	555	TTGCTCTCGCTGATGAGCGTTTCATCTGACCTGTCAAGCTTGCGCTCCGGGATGCTC	614
Dp	624	CTCCTCTCGCTGTGTGCGCTGCACTGCGCGGCAGTCCGTGGGACGAGGGAGTGTCTC	683
Qy	615	CAGGCGCTTGTGTGTGGGTATCCGTATGTTGTGTGCGGACACCTCTGTGTGCTCTC	674
Dp	684	CTGGGCGCTGTGCGGGGCGCGCACGGTATGTGCCACGCGACCGCTTCATCGATTCAG	743
Qy	675	GAGACCGCGGAGCTGAGCGCAGGGAAGGTGCGCTGCAAGGACATTTTCGGCGAGCGCGATG	734
Dp	744	CGGCGAGCGGGGCTGTGCCCCCGACGCGCGCTGTCAAGTCTTCTTCGCGACCGCGCACGG	803
Qy	735	GTTTCGACGAGCGAAGGGTGCCTGTGTGTTCTCAAGCGGCTCACTGAGGCCCGCGC	794
Dp	804	CACCACTGTGTCGAGGGGCGGAGCGGTGTGCTGTGCGCGGCTTCGACGCGCCGCG	863
Qy	795	GACGCGCATCGGATATTTGGCGGTGATTTGAGAGATCCGCATCAATACACGCGGTGAG	854
Dp	864	CTGTGGCTACCCCGTGCACGCGGTATCTCGGGGCGCGCGCTTCAACACGAGCGCGCGAG	923

QY	855	CAGGCGTCTGACCGTCCGCGGAACGGGAGGCTCCCAAGAAATCGTGCGTGAACAGGGGCGCTTGGC	914
Db	924	CGCGGGCTTGACCCGGGCCCAACGGACCGGGCGCAACAGCGGGGTGATCCGGCAGGACCTTGGC	983
QY	915	GGAGCGAAGGCTGCGCGCGCGTCTTTCGCTGGGTTTATGTCCAGGCGACACCGGACCGGACGAC	974
Db	984	CAAGCGACCGGTGACCGGCCGACACAGCGTGTGACCGGGTGTGAGGACACACCGGACCGGACCCC	1043
QY	975	GCTTGTGTATCCCATTCGAAATCCAAAGCTCTGAATGCGGTATACGGGCTTCGGGCGAGATGT	1033
Db	1044	GCTGGGCGACCCGATTCGAGGCCCGACGGCCCTCTCGGCGACCTACGGGCGGGGCCCGCGCGA	1103
QY	1035	CGCCACCGCGCGTCTGATCGGGTGGGTGAAACAACAACCTTGGCCATCCTGAGTATGCGTC	1093
Db	1104	GGGACGCGCGGTGTGGCTGGGCTGTGTAAATCGAACTCTGGGCCACACCCAGTCCGCGGC	1163
QY	1095	GGGGATCACTGAGGCTGCTGAAGAGTTCGTCTTCTTCCTTCAGACGGGCGAGATTCCTGCGCA	1154
Db	1164	CGCGCGCGGGCGGGCGGTCAATCAAGATGTGTATGTGGCCATGTGGGACCGGAGACGTGCCCCGAC	1223
QY	1155	CCTCCACGCGGAGCGCTGAAACCCCGGATCTCATGGGGTGATCTTCGGCTGACCGTTCAC	1214
Db	1224	GCTGACCTTCACGGAGCCACCCCGCGGCTGACATGTCGCGCGGTGACGTACGGGCTCT	1283
QY	1215	CGCGCGCCCGGACACCGTGTGGCCGGAATACGCCCGCGACCGGCGGGGGGTGAGCTCGTT	1274
Db	1284	GACCGAAGCGCCAGGACGTGGCCGGACACCGGACACACCGCGCGGTGCGCGGTCTGTCTCTT	1343
QY	1275	CGGCATGAGCGGAGCCAAACGCGGACGTTGTGTCTGTGAAGAAGGCGCGCGCGCGACGTTGCAC	1334
Db	1344	CGGCTTCAGCGGCGACCAACGCCCATGTGTATCTTGGAGGGCCCCGCCCGCGAGAGGACAC	1403
QY	1335	ACCGCG-----GGCGCGAGCGAACCGGACAGCTGTGTGTCTGTC	1376
Db	1404	GGAAGCGCGCGCTGCGCGGACGTCCTCGTGGACCGCGGGGCGCGGCTGCGGTGTCTTC	1463
QY	1377	GGCAAGACCGCGTTCAGCCTTGATGCAACAGGCGCGCGGCTGTGGCGACCATCTTGGAGAC	1436
Db	1464	CGGCGCGACGAGCGCGCGCTGTCCGAGCGGACGAGCGCGCTTGGCGGCCCATCTGACCGC	1523
QY	1437	CTACCGTTCCGAGTGTCTGGGCGGATGTGGGCGTTCACTGTGGCGACGACCGGACGCGCAT	1496
Db	1524	GCGCCCGGACCTTGGACACGGGCCGACGTGGCCACCGCGTGGCACACAGCGGGGCGGCTT	1583
QY	1497	GGAGCACCGGCTCGCGGTGCGCGGACGTCGAGGGAAGGGCGTGGCGGCGGCGCTTCGACGC	1556
Db	1584	CGACCAACGGGCGCGCGCTGTCCGAGCGGGAACCGTAGGAACTGTGTCGCGCGGCTTCGCGGC	1643
QY	1557	TGCGGCGGAGGACAGAGTGTGCGCGGTGCGGTGCGGTGCGAGTATCGGCCATTCCTACGCG	1616
Db	1644	CTTGGCGCACCGGAACCGCGCGCGCGCGCTGTGTACCGGCGGACCCCGCGCTCCGCGG	1703
QY	1617	CAACGTGCGCTTCTTCTTCAACCGGACAGGGGCGGACGACGCTGGGCAATGGGCGGTGGGCT	1676
Db	1704	CAAGGCGCGCTTCTTCTTCAACCGGACAGGGGACGACGACGCCGACATGGGCGCGGAAT	1763
QY	1677	GTACGATGTATGTTCGCGGTTCCCGGAGGCGCTTTCGACTGTGCGTGAAGCGCTTCAACA	1736
Db	1764	GGCGGCTTCACGACACCGTGTTCGCGGAGCGGCTTGGAGAGGTCTTGGCGCCAGCTGCACG	1823
QY	1737	GGAGCTGCAGCGGCGGCTTCGCGGAGGTGATGTGGGCGGAACCGGCGACGCTGCAGCGCGC	1796
Db	1824	GCACTCTGACCGGCGCGCTGCGCGAGGTGTGTTGCGCGGACGGGACGCGCCCGAGGCGCG	1883
QY	1797	GCTGCTGACCGACAGCGTTTCAACCGCGCGGCGGTGTACCTTTCGATATGGGCTTCGC	1856
Db	1884	CTGTCTGACACGACGCGCTTACACCGACCCCGCGCTGTTCGCGCTGAGGTTCGCGGCTCT	1943
QY	1857	CGCGCTGTGGCGGCTGTGGGTGTATAGACGCGAATTTGGTTCGCGCGGACATAGCATCTGTGA	1916
Db	1944	GCGGCTGTGGAGACTGGGCTTGGCGGCCCGGACATGTTCGCGGGGACACTCGGTTCGGGGA	2003
QY	1917	GCTGTGGCTGCTGCGTGGCGGCGGTGTCTTCGCTTGAAGACGCGGCTTTCGTGTGGC	1976



Db	2004	ACTGACCGCCGCTCACCAGCCGCCGGGGGTGTGGTTCGTCGCGCCACGCGCTGAGCGCCCTGTGTGCG	2063
QY	1977	TGCGCGCGGGCGCCTTGATGCAGGCGCTGCGCGCGCGCGGGCGATGTGTGTGATGCAGGC	2036
Db	2064	CGCCCGCGCGCGGTGACCCAGGCACTGCCCGCGGGCGGGCCGATGTGTGCGCGTGCAGGC	2123
QY	2037	GCCGAGGCGCGATGTGGCTGCTGCGGATGGCGCCGACGCAAGCTGTGATGTGCGCCG	2096
Db	2124	GACCGAGGACGAGGTGCGCGCCCACTCGCCGACGCGCCCGCGGTGGACATCTGGCGCG	2183
QY	2097	GGTCAAGCTCCGAGCCAGGTGTCTATCGCGGGCGCGGGGACACCCGTGCATGTGCATGCG	2156
Db	2184	CGTCAACGAGACCGGAGAGCGGTGTGTCTGTGCGGCACAGAGCCGCGGTCAAGCAGCTGGC	2243
QY	2157	GCGCGCATGTGCCCCGCGCGCGGGGCGCGAACAGAGGCGTTCACAGCTTCGATGTGCTTCA	2216
Db	2244	GCGGAGTGTGGCCCGCCGCGCGCGGAGAACAGAGGCTTGCGGGTCAGCCACGCGCTTCA	2303
QY	2217	CTCACCGCTCAATGAGCCCCGATGTGTGAGGCGCTTGCGGCGTGTGGCGGAGTCCGTGAGCTA	2276
Db	2304	CTCGGCCACCTTGAGCGCATAGACGAGGCGTTGCGCGAGGTTCGACGAGGGGTGTCTTA	2363
QY	2277	CCGCGCGGCGGTGATCGTCTGTGTGAGCAATCTGAGCGGAGAGGCTTGCACAGACGAGT	2336
Db	2364	CAGCGCGCGCTCCCTCCGCGGTGTCTTCACGCTCACCGGGCCCCCGTCACCGACGAGCT	2423
QY	2337	GAGCTCGCGGGGCTATTGGGTGGCCGACGCCGAGAGGTGTGTGCGCTTCGCGGATGAGT	2396
Db	2424	CCGAGAGCGGGAGACATGGGTGTGCGGCAGTCCGGAGACGGTGTGGCTTCCACGACGCGGT	2483
QY	2397	GAAAGCGCTGACACCGGCGCGGTGCGGACCTTGTCGAGGTCGGTCCGAATGACGCT	2456
Db	2484	CCGCGCGCTTGCGGCAACCGCGGGGCAACGCGCTTCTGTGAGGTTCGAGCCCGCGCGGTCT	2543
QY	2457	GCTTGAGCGCTGTGTGCTGTGCTGCATGTCCGAGCGCCGCGC---GGCGCTGTCCGATCGTC	2513
Db	2544	GAGGCGCGGCGCAAGCGGATGCTGTGCCGAGCGCGCCCCCGAGAGTTGTGTCCCGTCT	2603
QY	2514	GCGCGCTGTGGCTGACGAGCCGCGGACCTGTCTCGAGGCGCTTCGCGGGGCTTGTGGGCGGT	2573
Db	2604	GCGCGCGCGCGAGGCCCGAACCTGAGTCCGTCTGACGCGCCGTGCGGCGAGGCCACACGAT	2663
QY	2574	CGGTGGCTGTGTCTCTGTGGGCGCGGCTTTCCTTCAAGGGGGCGGCGGGTGTCCGCTGCC	2633
Db	2664	CGGCGCTTCGCGCGGCGTGGGACCGCGTCTCTCCAGAGCCCGGACCGCGCGTGCATTCGCC	2723
QY	2634	CACGTACCGCTTGGCAGGCGGAGCGCTACTGTGATGCAGACGAAAGCGC	2680
Db	2724	CAGGTAGCGCTTCCAGCGCGGCGCACTACTGTGCTGGCGGGCATGGCGC	2770

RESULT 3  
US-10-267-255-19  
Sequence 19, Application US/10267255  
Publication NO. US20030124689A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D  
APPLICANT: Mao, Y  
APPLICANT: Varoglu, M  
APPLICANT: He, M  
APPLICANT: Sheldon, P  
TITLE OF INVENTION: Micromycin biosynthetic gene cluster  
FILE REFERENCE: 600,456US1  
CURRENT APPLICATION NUMBER: US/10/267,255  
CURRENT FILING DATE: 2002-10-09  
PRIOR APPLICATION NUMBER: US 09/266,965  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: US 08/624,447  
PRIOR FILING DATE: 1996-08-19  
PRIOR APPLICATION NUMBER: PCT/US94/11279  
PRIOR FILING DATE: 1994-10-06  
PRIOR APPLICATION NUMBER: US 08/133,963

```

: PRIOR FILING DATE: 1993-10-07
: NUMBER OF SEQ ID NOS: 145
: SOFTWARE: FASTSEQ for Windows Version 3.0.
: SEQ ID NO 19
: LENGTH: 3978
: TYPE: DNA
: ORGANISM: Streptomyces lavendulae
: US-10-267-255-19

```

Query Match	19.1%;	Score 814.2;	DB 14;	Length 3978;
Best Local Similarity	57.6%;	Pred. No. 7.7e-170;		
Matches 1549;	Conservative	0;	Mismatches 1108;	Indels 30;
				Gaps 4;

OY	18	CGAGCGCGCAGCGCAGAAATCCGAAATTCGATTCGTCGAGACGAGTTGCGCTCGTCGCGCGTGG	77
Db	90	CGAGGAGCGCGGCCCCGCGAGCCCGTCGCGGATTCGTGGGCGATGCGCGTCCGCTTCCCCGGGGGA	149
OY	78	CGTATCGATTCGACCGGGATTCTTGGACGCTTCTGAGGGCTCGCGCGACACCGTCGGAGC	137
Db	150	CGTGGCATTCGCCCGGACGACCTGTGGCAGCTGTTCGCGAGGGCCGGGACCGCTGTACCGGA	209
OY	138	AGTCCCGCGCGGAAC---GCTGGGATGAGAGCGGTGGTTTGATATCCGACCCCGATGCGCCC	194
Db	210	GTTTCCCGCGGACCTGGGGCTGGGAGCTGACGCGCTTACGACCCCGAGCTGGGACATCC	269
OY	195	GGGGAGACGCGCCCGTTACGCGCGCATCTTTCTTGAAGCAGCGTACCGTCTTTCAGCGCTC	254
Db	270	GGGCAGAGACGATGACGCGCGCCACGCGCGGCTTCTCAAGAGACGCGCGCGGATTGCACGCGCG	329
OY	255	CTTCTTTGGGCGATCTCGCTCTCGCCAGAGGCGTGGCGATGAGACCTTCGACATCGACTCTTGCT	314
Db	330	CTTCTTTGGGATCATCGCGCGCGCAGGGGCTTCGCCATGACCCCGAGAGCGCGATGATCAT	389
OY	315	GGAAGTGTGCTGGAGGCGCTGAGAGACGCGCGCATGCTCTCATCGGCGCTGTCGGTAC	374
Db	390	GGAAGTCTCTCGGAGGCGCTTTCAGACAGCGGGGCGCTCGACCGACACACCTCGCGGGGGA	449
OY	375	GGAACCGGAGATGTTTCATCGGGATCGGCGCGCTCCGAATATGAGGCGCGCGTCCGCGAAGC	434
Db	450	GGAAGTCGGGCGTCTTCGTCGAGCTTCAACACACACGATCCTGATCAACGTCGTCGACGC	509
OY	435	GAGGGCGCTCGGAGAGATCGACGCTATGCGGGGCTGGGAGACATGCGCCAGCGTCGGAGC	494
Db	510	GCGGAGCGTGCAGGAGGCTTC-----ATGGGACCGGCACTTCGCGACACATCTCTC	563
OY	495	GGGCGCGAATCTGTATGCGCCCTCGGGCTGCGAGGGCGTGTGTGCGCGGTGATACGGGCTA	554
Db	564	CGGCGCGGTGCGCTACACCTTTGGCTTTCGAGGGCGCGGCGGTGTCCGTGCACCGGCTG	623
OY	555	TTGCTCTCGCTGTGTGGCGCTTTCATCTGGCTGTCAAGCTTGTGGGCTTCGGGGAAATGCTC	614
Db	624	CTCTCTCTCGCTGTGTGCGCTGCACCTGGGCGCGCACTCTCTGCGGAGGGGAGTCTTC	683
OY	615	CACGGCCCTGAGTGTGGGGTATGCTGATGATTTTCGCGGAGCACCTCGTGTGGTCTC	674
Db	684	CTTGGCCCTGCGGGGCGGGCGCGCGACGATGATGCGACGCGCATCGGCTTTCATCGAGTTTCAAG	743
OY	675	GAAAGACCCGGGCGCTGCGCAGGAGCGGTGCTGTCAAGGCAATTTTCGCGGAGGCGGATGG	734
Db	744	CCGCGACACGGGGGCTCGGCGCCCGACCGGCGCTGTGCAATCTCTTCGCGCGACCGCGACGG	803
OY	735	GTTTTCGACAGAGGGAAGGTGCGCGCTGTGTCTTCAACCGGCTCAGTGGAGGCGCGCG	794
Db	804	CACCACTGTGTGAGGGGCGGGCGCGGCTGTGTCTGTGCGCCGCGCTTCGAGGCGCGCGCG	863
OY	795	GGAACGGCGATTCGATATTTGGCGGTGATTCGAGATTCGCGATCATATCAGACGGTGGAG	854
Db	864	CTTGGGCTTACCCCGTGCAGCGGGTCAATCGGGGGGAGGCGGCTTCAACAGGACGGCGGAG	923
OY	855	CAGCGGTCTGACGCTGCGCAACGGGAGCTCCCAAGAAATGTGTCTGAACCGGCGCTTGGC	914
Db	924	CGCGGGCTTACCGCGCCCAACGAGACCGGGGCAACAGCGGGGTATATTCGGGACGAGCATCTGGC	983

QY 915 GGAGCAGGCTGGCGCGCTCTTGGTGGATTATGTCAGAGCAACGAGCGGAGCAAC 974  
Db 994 CAACGACGAGCTGACGAGCCGACAGCGTCAACCGGTCAGAGCAACGAGCGGAGCAACCC 1043  
QY 975 GCTTGTGACCCCATGAAATCAGACTTGAATGGGTATACCGGCTTCGAGCGAGATGT 1034  
Db 1044 GCTGGGCGACCCGATGAGGCGCCAGGCGCTCTCGCACCTACGAGGCGGCGCGGAGA 1103  
QY 1035 CGCCACCGCGCTGCTGATGGGTGGTGAAGACCAACTTGGCACTCTGATATGGCTG 1094  
Db 1104 GGGCAGGCGCTGTGGGTCTGCTGAAGTCGAACCTGGGCGCACCCCACTCCGCGCG 1163  
QY 1095 GGGGATCACTGGGCTGTGAAGGTGTCTTGTCCCTTGAGCAGGCGAGATTCCTGGCA 1154  
Db 1184 CGGCGCGGAGCGGCTGATCAATGATGTATGATGCGGACAGGAGCGCTCCCGCAC 1223  
QY 1155 CTTCCACGCGAGCGCGCTGAAACCCCGATCTCATAGGGGTGATCTTTCGCTGACGTCAC 1214  
Db 1224 GCTGACCTCAAGAGGCCACCCGCGGCTGACTGTGTCGCGGAGATGATCGGCTGT 1283  
QY 1215 GCGCGCGGAGACCGGTGGCGGACTGCAATACCGCGGAGCGGCGGAGGTAGCTGTT 1274  
Db 1284 GACCCAGGCGCAGAGCTGGCCCGGACACCGGACAGCGCGCGCTGGCGGCTCTGCTCTT 1343  
QY 1275 CGGCAATGAGCGGAGCAACGCGACGCTGGTGTGAAGAGGAGCGGCGGCGGAGCGTAC 1334  
Db 1344 CGGCTCAAGCGGACCAACGCGCCATGTATCTTGAAGGCGCGCGCGCGGAGAGGACAC 1403  
QY 1335 ACCGCG-----GCGCGGAGCGACCGGACAGCTGCTGTCTGTC 1376  
Db 1404 GAGCGCGCGCTGCGGAGCTCTCTCGGACGCGGCGGCGCGGCGGCTGTGGGTGTCTC 1463  
QY 1377 GGCAGAGACCGGCTGACGCTTGAATGACAGCGCGCGGCTGGCGGACATCTGAGAC 1436  
Db 1464 CGGCGCGAGCGGCGGCGCTGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1523  
QY 1437 CTACCTTCGAGTGTGTGGGAGATGTGGCGTTCAGTCTGGGAGCGAGCGGAGCGGAT 1496  
Db 1524 GGGCGCGACCTGCGACCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTT 1583  
QY 1497 GAGAGACCGGCTGCGGCTGGCGGCGGACGTCAGAGGAGGCGGCTGGGCGGCGGCGG 1556  
Db 1584 CGACACCGGCGGCGGCGGCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1643  
QY 1557 TGCGGCGGAGGAGACAGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1616  
Db 1644 CTTGGCACCGGAGCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1703  
QY 1617 CAAGCTCGCTTCTCTTCAACGAGCAGGCGGCGGCGGAGCGCTGGGAGTGGCGGCT 1676  
Db 1704 CAAGGCGGCTTCTCTTCAACGAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGACT 1763  
QY 1677 GTACGATGTATGTTCGCGCTTCGCGAGGCGGCTTCACTGTGCGGAGCTTCAACCA 1736  
Db 1764 GCGGCGTCAAGACAGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1823  
QY 1737 GAGAGCTCAACCGGCGGCTTCGCGAGGATGATGTGGCGGCGGCGGCGGCGGCGGCGG 1796  
Db 1824 GACCTTCGACCGGCGGCGGCGGAGGTGTGTTCGCGGCGGCGGCGGCGGCGGCGGCGG 1883  
QY 1797 GCTGCTCAACGAGCCTTCAACGAGCGGCGGCTTCACTTCAATATGCGGCTGCG 1856  
Db 1884 CTTGCTCAACGAGCGGCTTCAACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1943  
QY 1857 CGCGCTGTGGCGGTGTGTAGAGCGGAGGTTGTGCGCGGCGGCGGCGGCGGCGGCGG 1916  
Db 1944 GCGGCTGTGAGGAGCTGGGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2003  
QY 1917 GCTGTGTGCTGCTGTGTGTGGCGGCGGCTGTCTGCTTGAAGACGCGGCTTCTGTGCG 1976  
Db 2004 ACTGACCGCGGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2063  
QY 1977 TGCGGCGGCGGCGGCTGTGATGAGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2036

Db 2064 CGCCGCGGCGGCGGCTGACCCAGGCACTGCCCGCGGCGGCGGCGGCGGCGGCGGCGG 2123  
QY 2037 GCGGAGGCGGATGTGCTGTGGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2096  
Db 2124 GACCGAGGACGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2183  
QY 2097 GGTCAAGCTTCCGACGAGGTGTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2156  
Db 2184 GGTCAAGGAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2243  
QY 2157 GCGGCGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2216  
Db 2244 GCGGAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2303  
QY 2217 CTACCGCTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2276  
Db 2304 CTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2363  
QY 2277 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2336  
Db 2364 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2423  
QY 2337 GAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2396  
Db 2424 CGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2483  
QY 2397 GAGGCGGCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2456  
Db 2484 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2543  
QY 2457 GCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2513  
Db 2544 GACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2603  
QY 2514 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2573  
Db 2604 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2663  
QY 2574 CGGTGGCTGTCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2633  
Db 2664 CGGCGCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2723  
QY 2664 CACGTAACCTTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2680  
Db 2724 CACGTAACCTTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2770

RESULT 4  
US-09-953-348-74  
; Sequence 74, Application US/09953348  
; Publication No. US20030134398A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, David. H  
; APPLICANT: Mao, Yingling  
; APPLICANT: Varoglu, Mustafa  
; APPLICANT: He, Min  
; APPLICANT: Sheldon, Paul  
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER  
; FILE REFERENCE: 600.530US1  
; CURRENT APPLICATION NUMBER: US/09/953,348  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: PCT/US00/06394  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 153  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 74  
; LENGTH: 12249  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-953-348-74

Query Match 19.1%; Score 814.2; DB 12; Length 12249;  
Best Local Similarity 57.6%; Pred. No. 7e-170;  
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

Qy 18 CGAGCGCGACCGAAGATCCGATTGCGATGTCGAGCCGAGTTCCGTCGCCGATG 77  
Db 1077 CGAAGACCGCGCCCGCGAGCCCGTCGATCGTGGGACGTGGCGTCCCTTCCCGGGGA 1136  
Qy 78 CGTATGATGATGACCGGGTCTGGAAGCTCTCGAGGGCTCGCGGACACCGTGGGCG 137  
Db 1137 CGTGCATCGCCGAGACGACTGTGACGTGTGCTCCGAGGAGCGCGGACCGTACCGA 1196  
Qy 138 AGTCCCGCCGACAC---GCTGGATGACAGCGGTGTTGATCCCGACCCGATGCC 194  
Db 1197 GTTCCCGCCGACCGGGGCTGGAGGTGACGCGGTCTACAGCCCGGAGCGGGACCCC 1256  
Qy 195 GGGGAAAGCCCGCTTACGCGGATCTTTCTGAGCGACGTAGCTGTCTTGACGCTC 254  
Db 1257 GGGGAGGACGTACGCGCGCACGCGGCTTCTTCAAGGACCGCGCGGATTCGACGCGC 1316  
Qy 255 CTTCGTGGGATCTTCGCTCGGAGGCGTGGATGAGCCGTGACATGACATCTTCT 314  
Db 1317 CTTCGTGGGATCACGCGCGGAGGCGCTGCGCATGAGACCGGACGCGCATGATCAT 1376  
Qy 315 GAGAGTGTGAGGAGCGCTGAGAGACCGCGCATGCTCCATGCGGCTGTGCGTAC 374  
Db 1377 GAGAGTCTCTGGAGGCGTTCGAGCAGGCGGCTCCAGCGACCACTCTGCGGGGGA 1436  
Qy 375 GAAAAGGGAGTGTTCATCGGAGTGGCCGCTCGAATATAGAGCGCGCTGCGGAGC 434  
Db 1437 GAGAGTGGGCTCTTCGCTCGGCTCCAGAGCAAGACTGATCAAGTGTGCTGAGC 1496  
Qy 435 GAGCGCTCGGACAGATCGACGCTCAGGCGGCTGGAGAGCATGCCACGCTGAGC 494  
Db 1497 GCGGACGTCCGCGAGGCTTC-----ATCGGAGCGGCAACTCCGCGACATCTCTC 1550  
Qy 495 GGGCGGATCTGATGCTCTCGGCTGCGAGGCGGTGTGCGGAGTACGCGCTA 554  
Db 1551 CGGCGGCTGCGCTACCTTCGGCTTGAAGGCGCGCGCTGTCTCGTACACGCGCTG 1610  
Qy 555 TTCTCTCTGCTGTGCGCTTTCATCTGCGCTGTCAGAGCTTGGCGCTCGGGAAATGCTC 614  
Db 1611 CTCTCTCTGCTGTGCGCTGTCGACCTGCGCGAGTCTCTGCGGAGGAGTGTCTC 1670  
Qy 615 CAGGCGCTGCTGCTGGGTATGCTGATGTCGCGAGCACCTCTGTGTGCTCTC 674  
Db 1671 CTTGCGCTGCGCGCGCGCGAGCGGTATGCGCACGCGCTTTCATGATTCAG 1730  
Qy 675 GAAAGCCCGGCGCTGCGACAGGAGCGTCTGTCAGAGCATTTTGGCGGAGCGGATG 734  
Db 1731 CGGCGAGGCGGCTGTGCGCGCGCGCGCGCTGTCAGAGTCTTCTGCGGACCGCGACG 1790  
Qy 735 GTTTCGACGAGGCGAAGGCTGCGCGCTGTGTCTCAAGCGGCTCACTGAGACCGCGC 794  
Db 1791 CACCACTGTGTCGAGGCGCGCGCGCTGTGCTGCGCGCTCTCGAGCGCGCGC 1850  
Qy 795 GAGCGGATGAGATTTGGGCTGATTCGAGGATCCCGGATCATACAGCGGTGCGAG 854  
Db 1851 CTTGCGCTACCCCGCTGACGCGGTATCCGCGGACGCGCTTCAACGAGACGCGGCTAG 1910  
Qy 855 CAGCGTGTGACCGTGCAGAGGAGGCTCCCAAGAAATCGTCTGAAACGCGGCTGCG 914  
Db 1911 CGCGGCTGTACCGCGCGCGCGCGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 1970  
Qy 915 GAGCGAGGCTGCGCGCGCTCTTCTGAGGATTTATGTCAGGACACGCGACGCGGACGAC 974  
Db 1971 CAAGCGAGGCTGACGCGCGCGAGCGCGGTGAGCGGACACGCGGACCGGACCGC 2030  
Qy 975 GCTTGTGACCCCGATTCGAAATTCAGCTGTGATTCGCGCTTCGCGGAGATGT 1034  
Db 2031 GCTGGGACCCCGATTCGAGGCGCGCGCTCTCTCGCACCTACGCGGCGCGCGCGGAG 2090

Qy 1035 CGCCACCGCGCTGCTGATCGGCTCGGTGAGACCAACTTGGCCATCTGATGCTC 1094  
Db 2091 GGGGAGCGCGCTGTGCTGCGGCTCGTGAAGTCAACTTGGGCGACACCGAGTCCGCGC 2150  
Qy 1095 GGGGATCACTGGGTGCTGAAGGTGTCTTCTCCCTTACAGACGGGAGATTCCTGCGCA 1154  
Db 2151 CGGCGGCGCGCTGTATCAAGATGTGATGCGCATCGGACGAGGAGCTGCGCCGAGC 2210  
Qy 1155 CTTCCAGCGCGAGCGCTGAACCCCGGATTCATGAGGAGTATCTTTCGCTGACCGTAC 1214  
Db 2211 GCTGCACTACAGAGGACCGACCCCGCGCTGACTGTGCTGCGGTACGTAAGCTGCT 2270  
Qy 1215 GCGCGCCCGACACCGTGGCGGACTGGAATACGCGCGACGCGGCGGAGTGAAGCTGCT 1274  
Db 2271 GACCGAGGCCAGAGACTGGCGGACCGGACACCGGACCGCGCTGCGCGCTGTCTCT 2330  
Qy 1275 CGGATAGCGGAGACCAAGCGGACGCTGTGCTGGAAGAGGCGCGCGGAGAGTGCAC 1334  
Db 2331 CGGCTAGCGGACCAAGCGGACGCTGTGATCTGGAAGGCGCGCGCGCGGAGGAGC 2390  
Qy 1335 ACCGCG-----GCGCGGAGCGACCGGACAGCTGTGCTGTCTC 1376  
Db 2391 GAGCGCGCGCTGCGGACGCTCTCTGACAGCGCGGAGCGCGCTGCGGTGCTCTC 2450  
Qy 1377 GGAAGGACCGCGTCAAGCCCTGATGACAGCGCGCGCGCTGCGGACCATCTGAGAC 1436  
Db 2451 CGGCGGAGGAGGCGCGCTGCGAGCGGAGGCGGAGGCGCTGCGGCGGACCTGACGCG 2510  
Qy 1437 CTACCTTTCGAGTGTCTGGGCGATGTGCGCTTCACTGTGCGACGACGCGGAGCGAT 1496  
Db 2511 GCGCGCGACCTTGCACCGCGCGGACGCTGCGCACCGCGCTGCGCACACGCGGCGCTT 2570  
Qy 1497 GAGACACCGGCTGCGGCTGCGGCGGACGCTGAGGAGGAGTGGGAGCGCTGAGCGC 1556  
Db 2571 GACACACCGGCGCGCGCTGCTGCGCGGAGCGGAGCGGAGTGTCTGCGCGCTGCGCGC 2630  
Qy 1557 TGCAGCGAGGACAGACGCTGCGCGGCTGCGGAGTATCGCGATTCCTACGCGG 1616  
Db 2631 CTTGCGCACCGGAAACCGCGCGCGCGCGCTGTGACCGGCGGAGCCCGCGCTCGGCGG 2690  
Qy 1617 CAACTGCGCTTCTTTCACCGGACGAGGAGGCGGACGCTGAGGATGGGCTGTGCT 1676  
Db 2691 CAAGCGCGCTTCTCTTTCACCGGACGAGGAGGACGACAGCGCGGATGGGCGCGGACT 2750  
Qy 1677 GTACGATGATGCTGCGCGCTTCCGCGAGGCGCTTGCAGCTGTGAGGCTGTCAACA 1736  
Db 2751 GCGCGCTCACAGACCGGTTCGCGGACCGCTTGAAGAGTGTGCGCGGACGCTGACCG 2810  
Qy 1737 GAGCTCGACCGCGCGCTCCGCGAGTATGTGGCGGAAACCGGACGAGCTGACGCGCG 1796  
Db 2811 GACCTCGACCGCGCGCTGCGCGGAGTGTGCTGCGCGGAGGAGGAGGAGGAGGAGG 2870  
Qy 1797 GCTGCTGACGACGAGCTTTCACCGGCGCGCTGTTACCTTTCGAAATGCGCTGCG 1856  
Db 2871 CTTGCTGACGACGAGCTTTCACCGGCGCGCTGTTGCTGCTGAGGCTGCGCTGCT 2930  
Qy 1857 CGGCTGAGCGCTGCGGCTGAGAGCGGAGTGTGCGCGGAGTATGATCGGAG 1916  
Db 2931 GCGGCTGAGGAGATGCGGCTTTCGCGCGCGGAGTGTGCGGAGCTTGTGAGGAG 2990  
Qy 1917 GCTGAGCTGCTGCTGCTGCGGCGGAGTGTCTGCTTTCGAGACCGGCTTCTGAGTGC 1976  
Db 2991 ACTGACCGCGCTTACCGCGCGGAGTGTGCTGCTGCGGAGCGCTGCGGCTGCTGCTG 3050  
Qy 1977 TGCAGCGGCGCTGATGACGCGCTGCGGCGCGGCGGAGTGTGATGAGGCG 2036  
Db 3051 CGCGCGCGCGCTGACCGGAGGAGTGTGCGCGCGGCGGAGTGTGCGCGGAGGCG 3110  
Qy 2037 GCGGAGGCGGATGCTGCTGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2096  
Db 3111 GACCGAGGAGGAGTGTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3170  
Qy 2097 GGTCAAGCTTCGAGACGAGTGTATCGCGGCGCGGAGGAGGAGGAGGAGGAGGAG 2156

Db	3171	CGTCAACGCAACCGGAAGCGGTGTGTCTCTCCGGCCACGAGGCCCGCCCTTCACGGAACCTGGC	3230
Qy	2157	GGCGGCGATGGCCCGCGCGCGGGGCGCGAACCAAGCGCTTCCAGTCTTCGCAATGCGTTCCA	2216
Db	3331	GGCGAGTGGGCGCGCCCGCGCGCGGAGAACGAGGCGTGGCGGGTCACACGCGCTTCCA	3290
Qy	2217	CTCAACCGCTATAGGCCCCGATGCTGGAGAGCGTTGGGCGGTGGGCGAGTGGGAGGCTA	2276
Db	3291	CTCCGCCCACTGGAGCGCATGACCGAGAGCGTTTCGCCAGAGTTCGACAGAGGGGTCTTA	3356
Qy	2277	CGGCGGCGCGTGCATGTCGTCTGTGTCAGCAATCTGAGCGGGAAGCTTGCACAGACGAGT	2338
Db	3351	CAGCGCGCGCTCCCTCCCGGTGTGTCTCAGCACTCACGGGGCCCCCTGCACGACGAGCT	3410
Qy	2337	GAGCTCGCGGGGCTATTGGGTGGCCCAAGCCCGAAGAGTGGTGGCTTTCGGGATGAGT	2396
Db	3411	CCGCAAGCGCGGAACACTGGGTGGCGGCAAGTCCGGGAGACGGGTGCGCTTTCACAGACGCGT	3470
Qy	2397	GAAGGCGCTGACACGCGCGGTGCGGAGCACTTCGTCCAGGTCGGTCCGAAATGACGCT	2456
Db	3471	CCGGGCCCTGGCGCAACGCGGGGCGACGCGCTTCCTGGAGGTCCGGGCGCGGCGGTGCT	3530
Qy	2457	GCTGGGCGCTGTGCTGCTGCTGCATGACCAGACGCCCCGCTC---GGCGCTCTGCATGCTC	2513
Db	3531	GACGGCGCGGCAACGCGCGCATCTGTGCCGAGCGCGCCCCCGAGACGTTGTGTCCTCCGTGCT	3590
Qy	2514	GCGCGCTGGGCGGTGACGAGCGGAGCAACGATGCTCGAAGCGCTCGGCGGGACTCTGGGACGT	2573
Db	3591	GGCGGCGCGGACGCGCCGAACTCCGATCCGTGTACAGGCGGTGCGGACGGCCACACGAT	3656
Qy	2574	CGGTGGCGTGTCTCTCGGAGCGCGGCTTCTCCCTCAGGGGGCGCGCGGGTGCCTGCG	2633
Db	3651	CGGCGCTTCGCGCGGCGTGGGACCGGCTGTCTCCCAAGCGCGGACGCGCGTGGACCTGC	3710
Qy	2634	CACGTAACCTTTGGACGCGCGGACCGCTATCGAATCGACACGAAAGCGG	2680
Db	3711	CACGTAGCCCTTCAGCGCGGCGCACTACTGTGCTGGCGGGCAATGGCGG	3757

```

RESULT 5
US-10-267-255-74
; Sequence 74, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Vargolu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600,456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 12249
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-74

```

Query Match	19.1%;	Score 814.2;	DB 14;	Length 12249;
Best Local Similarity	57.6%;	Pred. No. 7e-170;		
Matches 1549;	Conservative	0;	Mismatches 1108;	Indels 30; Gaps 4;

Oy	18	CGACGCGGAGACCGAAGATCCGATTTGGGATCGTGAAGCGAGTTGGCTCGCGCCGCGGG	77
Db	1077	CGAGAGCGCGGCCCGCGAGCCCGTCGCGAGTGTGGCATGGCGTGTCCGCTTCCCGGGGA	1136
Oy	78	CGTATCGATTCAGCGGGTTCTTGAACGCTTCGAGGGCTCGCGCGACACGTCGAGG	137
Db	1137	CGTGCATTCGCGCGAGACCATCTGTGGCAGCTGTGTGCGGAGGGCCGGGACGCGTCACCGA	1198
Oy	138	AGTCCCGCGCGAAC---GCTGGAGATGAGACGCGTGGTTTGAATCCGACCCCGATGCCCC	194
Db	1197	GTTCCCGCGACCGGGGCTGAGGACGTGACGCGCTTACGACCCCGAGCCGGGACCCC	1256
Oy	195	GGGAGACGCGCCGTTACGCGCGCATCTTTCCTGAGCGACGTACCGTCTTTCAGCGCTTC	254
Db	1257	GGGAGAGACGTAGCGCGGCCACGCGCGGCTTCTTCAGAGACGCGCGGATTTCAGCGCGC	1316
Oy	255	CTTCTTTGGGCATCTTCGCTCGCGAAGCGCTGCGGATGAGACCTTCACATCGACTCTTGTCT	314
Db	1317	CTTCTTTGGGCATCACGCGCGCGGCGCTTCGCGATGACCCCGAGACGCGCATGATCAT	1376
Oy	315	GGAGGTGCTGGAGGCGCTGAGAGAACGCGCGCATGCTCATCGCGCTCGTCCGTAC	374
Db	1377	GGAGGTCTCTGGAGGCGTTTCAGACAGCGGCGCTGACCGCACACCTCTCGGGGCGGA	1436
Oy	375	GGAACCGAGAGTGTTCATCGGAGTCGCGCCGTCGGAATATGAGGCGCGCGCTGCCGAAAG	434
Db	1437	GGAGTGGGCGCTCTTCGTCGCGCTTCMAAGACAGCATCTGTATCAACGTCTTCGACGC	1496
Oy	435	GACGCGCTCGCGAAGATTCAGCGCTCATGGCGGGCTGGGGAGATGCCACGCTCGAGC	494
Db	1497	GCGGAGCGTCGCGAGGGCTTC-----ATCGGAGCGGCACTTCGCGACATCTCTTC	1550
Oy	495	GGGCGGATCTCGATGACCCCTCGGCGCTGCGAGGCGCGTGTGTGCGGCGATTCAGCGCTA	554
Db	1551	CGGCGGCGGCTACACCTTCGCGCTTCAGAGGCGCGGCGGTGTTCGTGCACACGCGCTG	1610
Oy	555	TTGCTCTCGCTGTGCGCGCTTATCTGCGCTGTACAGCTTGGCGTTCGCGGGAATGCTC	614
Db	1611	CTCTCTCTCGTGTTCGCGCTGACCTGCGCGCGACATCTCTTCGGGAGGAGGTGCTC	1670
Oy	615	CACGCGCTCGGCTGTGAGGAGTATGCTGATGTTTCGCGAGCACCTCGTGTAGCTCTC	674
Db	1671	CTGGCGCTGGCGGCGCGCGACGATGAGCCACCGCACCGCTTTCATTCAGATTTCAG	1730
Oy	675	GAAAGCCCGGCGCTGCGCAAGGACGCGTCTGTCAGAGCATTTTCGCGGAGGCGCGATG	734
Db	1731	CCGCGACCGGGGCTGTGCGCCCGCGCGCGCTGCGATCTCTTCGCGCGACCGCGGACGG	1790
Oy	735	GTTTCGACGAGGCGGAAGGTCGCGCGTGTGTCTTCAGAGGCGCTTCAGTGAAGCCGCGC	794
Db	1791	CACCACTGTGTCCAGGCGCGGCGCTGTCTGTCTGTGGCCCGGCTTTCGACGCGCCGCG	1856
Oy	795	GAGCGGCGATCGATTTATTTGGCGATTCAGATCCGCGATCAATCAACGACGCTGCAG	854
Db	1851	CCTGGGCTAACCCGTGACACGCGGTCAATCCGGGCGAGCGCGTCAACACAGGCGCGCGAG	1910
Oy	855	CAGCGGTCTGACCGTTCGCGAAGCGGAGCTCCCAAGAAATGCTGTGAAGCGGGCGCTGGC	914
Db	1911	CGCGGCGCTTACGCGCGCCAAACGAGCGCGGCAACACGCGGTGTATCCGGACGGCATGGC	1970
Oy	915	GAGCGAGCGTGCAGCCCGCTTTCGATGGGTTATGTGAGGCAACAGGCAACGGGACGAC	974
Db	1971	CAGCGACACGCTGACGCGCACACGCTTCAGACGCGGTGAGGCAACACGGCACCGGACCCC	2030
Oy	975	GCTTGTGACCCATCGAATCCAGGCTCTGAATGCGGTATCGGCTTCGAGCGAGATGT	1034
Db	2031	GCTGGGCGACCGATTCAGGCGCCAGGCGCTTCCTGCGACCTACCGGCGGCGCGCGGA	2090
Oy	1035	CGCACCGCGCTGTGTATTCGGGTGCGGTGAAGAACCAACTTGGGCATCTGTGATTCGCTC	1094
Db	2091	GGGAGGCGCGGTGTGGCTGGGCTGTCTGATBAATCAACTTGGGCAACACCACTGCGCGC	2150

Db 2091 GGGCAGGCCGCTGTGGCTGGGCTCGCTAAGTGAACCTGGGCGACACCCAGTCGGCGC 2156

QY 1095 GGGGATCATGGGCTGCTGAAGTCTGTTGCTTCCCTTACGACCGGACGAGTTCTCGGCA 1154  
Db 2151 CGGGCGGGCGGCGCTCATCAAGATGTGTATGATGCGGACCGGAGCGTCCCGCAC 2210  
QY 1155 CTTCCAGCGGACGGGCGGTGAAGCCCGGATCTCAATGGGGTATCTTCCGTCGACCGTAC 1214  
Db 2211 GCTGACCTTACGGAGGCCACCGCGCGGTGACTGTCCCGGTGACGTACCGGCT 2270  
QY 1215 GCGCGCCCGGACACCGTGGCCGACCTGGAATACCGCGGACCGGCGGGGTAGCTGCT 1274  
Db 2271 GACCGAGGCCAGGACTGGCCGGAACCGGACCGCGCGCGTGGCGCGCTGCTGCTT 2330  
QY 1275 CGGATGAGCGGAGCAACCGGACGTTGTCTGAGAGAGCGCGCGGCGAGCGTGCAC 1334  
Db 2331 CGGGGTGACGGCGACCAACCGCCATGTATCTTGAAGGGCCCGCGCGCGGAGAGGAC 2390  
QY 1335 ACCGCG-----GCGCGGAGAGACCGGACGAGCTGTGTGTCTG 1376  
Db 2391 GAGCGCGCGCTGCCGAGCTCTCTCGACCGCGGGGCGCGCTGTGCGGTGTCTG 2450  
QY 1377 GGCAGAGACCGGCTCAGGCTGATGACAGGCGCGCGCTGGCGGACATCTGAGAC 1436  
Db 2451 CGGCGGAGCGGACGGCGGCTTCCAGCGGACGAGCGCGCTGGCGCGCTGACCGC 2510  
QY 1437 CTACCTTTCGAGTGTCTGGCGATGTGGCTTCAGTGCGGACGAGCGCGCGGAT 1496  
Db 2511 GCGCGCGGACCTTGGCACCGGCGGACGTTGCCACCGCGCTGGCACGCGGCGGCTT 2570  
QY 1497 GAGACACCGGCTCCCGGTGGGCGGACGTCAGAGGAGGGGCTGGGCGGACCGCTGAGAC 1556  
Db 2571 CGACACCGGCGCGCGCTGTGTGGCGGAGCGTGAAGAACCTGTGCGCGCGCTGCGCG 2630  
QY 1557 TGGCGCGGAGGAGACAGTGTGCGCGGTGGCGGACGATGCGGACATCTTCAACGCG 1616  
Db 2631 CTTGGCGGAGGAGACCGCGCGCGCGCGCTGTGTGACCGCGGACCGCGCGCTGCGCG 2690  
QY 1617 CAAGCTGCGCTTCTTCTTACCGGACAGGAGGCGGACGCTGGGATGGGCGCT 1676  
Db 2691 CAAGGCGCGCTTCTTCTTACCGGACAGGAGGCGGACGCGGATGGGCGGACCT 2750  
QY 1677 GTAGATGTATGTGTCCGCTTCCGCGAGGCGTTCGACCTGTGTGGAGGCTGTTCACCA 1736  
Db 2751 GCGCGCTTCACAGCACCGGTGTGCGGAGCGCTTGAACAGGTGTGCGCGGCTGACCG 2810  
QY 1737 GGAAGCTCGACCGCGCGCTCCGCGAGGTATGTGGCGGACCGGACGCTGACGCGCG 1796  
Db 2811 GCACTCTGACCGCGCGCTCCGCGAGGTCTGTGCGCGGACGCGGACCGCGGCGCG 2870  
QY 1797 GCTGCTGACGACAGCTTTCACCGGACCGCGGCTGTTCCTTGAATATGGCTGCG 1856  
Db 2871 CTTGCTGACGACAGCGGCTTACCGGACCGCGGCTGTTCGCGTGAAGTGTGCGCT 2930  
QY 1857 GCGCGTGTGGGCTGTGGGTGTGAAGCGGAGTGTGCGCGGACATGACATCGTGA 1916  
Db 2931 GCGGCTGTGAGAGACTGTGGGCTTGGCGCGCGGACGATGTTCGCGGCTCACTGCTG 2990  
QY 1917 GCTGCTGCTGCTGCGGTGGCGGCGGTGTTCGCTTGAAGACGCGGCTTCTGTTGCG 1976  
Db 2991 ACTGACCGCGCTTACCGCGCGGCGGTGTGCTGCTTCCCGACGCTGCGCGCTGTGCG 3050  
QY 1977 TGCAGCGGCGGCTGTATGACGCGCTGCGCGCGCGGCGGAGTGTGTGATGAGCG 2036  
Db 3051 CGCGCGCGGCGGCTGACCGGACGCTGCGCGGCGGCGGCGGACATGTGTGCGCGGCG 3110  
QY 2037 GCGCGAGCGGATGT 2096  
Db 3111 GACCGAGGACGAGGT 3170  
QY 2097 GGTGACGCTGCGGACGAGT 2156  
Db 3171 GGTGACGAGGACGAGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3230  
QY 2157 GCGCGGATGTGCGCGCGCGGCGGCGGACCAAGGCGCTTCACTTTCGATGTGCTTCA 2216

Db 3231 GCGGATGTGAGCGCGCGCGCGCGCGGAGACAGAGGCTCGGGTCAAGCACCGCTTCA 3290  
QY 2217 CTCACGCTATGAGCGCGCGATGTGTGAGCGCTTGGGCGGTGTGCGGATGCTGAGCTA 2276  
Db 3291 CTCGCGGACCTGAGCGCGGATGTGTGAGCGCGCTTGGGCGGTGTGCGGATGCTGAGCTA 3350  
QY 2277 CCGGCGCGGCTGT 2336  
Db 3351 CAGCGCGCGGCTGT 3410  
QY 2337 GAGTGTGCTGAGT 2396  
Db 3411 CCGGAGCGCGGACACTGT 3470  
QY 2397 GAGGCGCTGACGCGCGCGGCGGAGCGGACCTTGTGTGTGTGTGTGTGTGTGTGTGT 2456  
Db 3471 CCGGCGCGGCTGT 3530  
QY 2457 GCTGCGGCTGT 2513  
Db 3531 GAGCGCGCGGACGCGGAGT 3590  
QY 2514 GCGCGCTGT 2573  
Db 3591 GCGCGCGGAGGCGCGGACCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3650  
QY 2574 CCGT 2633  
Db 3651 CCGCTTGT 3710  
QY 2634 CACGTACCTTGT 2680  
Db 3711 CAGCTACGCTTGT 3757

RESULT 6  
US-09-953-348-96  
Sequence 96, Application US/09953348  
Publication No. US20030134398A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, David. H  
APPLICANT: Mao, Yinqing  
APPLICANT: Vargolu, Mustafa  
APPLICANT: He, Min  
TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER  
FILE REFERENCE: 600.530US1  
CURRENT APPLICATION NUMBER: US/09/953.348  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: PCT/US00/06394  
PRIOR FILING DATE: 2000-03-10  
PRIOR APPLICATION NUMBER: 09/266965  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 153  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 96  
LENGTH: 18331  
TYPE: DNA  
ORGANISM: Streptomyces lavendulae  
US-09-953-348-96

Query Match 19.1%; Score 814.2; DB 12; Length 18331;  
Best Local Similarity 57.6%; Pred. No. 6.8e-170;  
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

QY 18 CGAGCGGCGGACCGGAGATTCGATGCTGCGAGCGAGTTCGCTGCGCGGTG 77  
Db 1077 CGAGGAGCGGCGCGGCGGAGCGCGGATGCTGCGAGCGAGTTCGCGCGGAG 1136  
QY 78 CGGTATGATCTGAGCGGCTTGTGAGAGCTCTCTGAGAGGCTCGGCGACACCTGTG 137  
Db 1137 CGTGTATGCTGCGGACGACTGTGTGCGAGTGTGTGCGAGGCGCGGACCGCTGACCGA 1196

QY 138 AGTCCCGCCGAC---GCTGGATGACAGACGCTGTTTATCCCGACCCCGATGCC 194  
DB 1197 GTTCCCGCCGACCGGGGCTGGAGCTGACGCGCTTACGACCCCGAGCCGGACCCC 1256  
QY 195 GGGGAGAGCGCCGTTTACGCGCGCATCTTCTTGAGCGACGATGCTGTTGAGCGCTC 254  
DB 1257 GGGGAGAGAGTACGGCGCGCACGGGGGCTTCTTCAAGAGACGCGCGGATTTGAGCGCC 1316  
QY 255 CTTCTTGGCATCTCGCTCGCGAAGCGCTGCGATGAGCCTTGACATCGACTCTTGT 314  
DB 1317 CTTCTTGGCATCAGCGCGCGAGCGCTCGCATGAGACCGGACGCGATGATCAT 1376  
QY 315 GGAAGTGTCTGGGAGCGCTGAGAGACCGCGCATGCTCATGAGCGCTGTGGTAC 374  
DB 1377 GAGGATCTCTGGAGGCGTTTGACAGCGGGCGCTTGACGAGCACCTTCGGGGCGA 1436  
QY 375 GGAAGCGGAGTTCATCGAGATCGGCGCTCGCATATGAGCGCGCGTCCGAGC 434  
DB 1437 GAGCGTGGGCTTCTGTGCGCTCAACAGCAACGATCTGATCATGCTCTGAGCG 1496  
QY 435 GACGCGTCCGAGAGATGACGCTCATGCGGGCTGGGAGCATGCCAGCGTGGAGC 494  
DB 1497 GCGGAGCGTCCGAGGCGCTTC-----ATCGGACCGGCACTCGGACATCTCTC 1550  
QY 495 GGGCGGAACTGTGATGCGCTGGGCTGGAGGCGCTGTCTCGCGGTGATACGGCTTA 554  
DB 1551 GGGCGCGTGGCTTACCTTGGGCTTGAGGGCGCGGCGTGTCTGTGACACCGCTG 1610  
QY 555 TTGCTCTGTGATGCGCTTCACTGAGCTGTGACAGCTTGGCTCGGAGGAATGCTC 614  
DB 1611 CTCTCTCTGTGTGCGCTGACCTGGCGCGGAGTCTTCCGCGAGGGGAGTGTCTC 1670  
QY 615 CACGCGCTGTGATGCGCTTATGCTGATGTTGTGCGGAGCACCTTGTTGTCTC 674  
DB 1671 CTGTGCGTGGCGGCGCGGCGACGATGATGACACGCGCATCATGATTCAG 1730  
QY 675 GAAAGACCGGCGCGCTGCGAGGAGCGTGTGAGAGGATTTTGGCGGAGCGCGATG 734  
DB 1731 CCGGACGCGGGGCTGGGCGCGCGGCGCGCTGCAAGTCTTCTGGGCGACCGCGAG 1790  
QY 735 GTTCCGAGCGAGGAGGATGCGCGTGTGCTTCAAGCGGCTCAGTGAAGCGCGCG 794  
DB 1791 CACCACTGTGTGAGGGGCGGCGGTGTGTGTGTGCGCGCGCTTCCGAGCGCGCG 1850  
QY 795 GAGCGCGATGCGATATTTGGCGGTATTTGAGAGATTCGCGATCAATCAGCGGTG 854  
DB 1851 CTTGGGCTAACCCGCTGACGCGGTATCGGGGCGAGCGCGTCAACAGAGAGCGGCG 1910  
QY 855 CAGCGGTGACCGTGGCGGAGCGGAGCTCCCAAGAAATCGTGAACCGGCGCTGCG 914  
DB 1911 CGCGGCGCTGACCGGCGCGCCAGAGCGGCGCAACGCGGTATTCGCGAGCACTGCG 1970  
QY 915 GAGCGAGGCTGCGCGCTTCTGGTGGTTATGTGAGGACACAGCGAGCGGACGAC 974  
DB 1971 CAACCGACGCGTGAAGCGGAGAGCGTCAACGCGGTGAGAGCAACGAGCGGACCCC 2030  
QY 975 GCTTGTGATCCCATGAAATCAAGCTTGAATGCGGTATACGCGCTTGGCGAGAT 1034  
DB 2031 GCTGGGCGACCGGATGAGAGCGCAAGCGCTCTCTCGCACATACGCGGCGGCG 2090  
QY 1035 CGCGACCGCGCTGATGCGGTGAGTGAAGACAACTTGGCGCATCTGATATGCTG 1094  
DB 2091 GGGGAGGCGGTGTGGCTGGGCTGTGAGGATGAACTTGGGCGACCACTGCGCGCG 2150  
QY 1095 GGGGATCACTGGGCTGTAAGTGTCTTGTCTTCAAGCGGAGAGATCTTGCGCA 1154  
DB 2151 CGGCGCGCGCGCTGATCAAGATGTGATGCGCATGCGGACAGCGAGCTGCCGCGAC 2210  
QY 1155 CTTCAACGCGGAGCGGCTGAACCCCGGATCTCATGGGTGATCTTTCGCTGACCGT 1214  
DB 2211 GCTGACCTCAACGAGGCGGCGCGGCTGATGCTGCGCGGATGAGTACGAGCTGCT 2270

QY 1215 GCGGCGCGGACACCGGTGCGCGGACTGGAATACCGCGGAGCGGGGCTGAGCTGTT 1274  
DB 2271 GACCGAGCCCAAGACTGGCCGAGACCGGACAGCGCGCGCTGTGCGCTGCTCTT 2330  
QY 1275 CGGATGAGCGGAGACCAACGCGCATGTGTCTGGAAGAGCGCGCGCGGACGCTGAC 1334  
DB 2331 CGGCGTCAAGCGGACCAACGCGCATGTGATCTGGAAGGCGCGCGCGGAGAGGAC 2390  
QY 1335 ACCGCG-----GCGCGGAGCGACCGGACAGACTGTGCTGTCTC 1376  
DB 2391 GAGCGCGCGCTGCGCGACGCTCTCTGACACCGCGGCGCGCTGCTGCTGTGCTC 2450  
QY 1377 GGAAGAGCGGCTGAGCGCTGATGACAGAGCGGCGCGGCTGCGCGCATCTGAGAC 1436  
DB 2451 CGGCGGACGAGGCGGCTGCTCGAGCGAGGCGAGCGCTTGGCGGCTTCAACCGC 2510  
QY 1437 CTACCTTTCGAGTGTGAGCGATGTGCGCTTACGCTGCGACGACGCGAGCGGAT 1496  
DB 2511 GCGCGCGACCTGGAGCGCGCGGACGTTGCGCACCGCGCGCTGCGACAGCGGGCGCTT 2570  
QY 1497 GAGGACCGGCTCGCGGTGCGCGGACGTCAGAGGAGGGGCTGCGGCGACCTGAGAC 1556  
DB 2571 CGACCAACGCGCGCGCTGTGCGCGGAGCGGTGAGGAACTGTGCGCGCGCTGCGGCG 2630  
QY 1557 TGCGGCGAGGAGACAGACTGCGCGCGGCGGTGCGAGTATCGCGCATCTTCAACGCG 1616  
DB 2631 CTTGCGACCGGAAACCGCGCGCGCGCTGCTGATCACGCGCGGACCCGCGCTCGGCG 2690  
QY 1617 CAAGCTCGCTTCTTCTTCAACCGGACAGAGGAGCGGACGCTGAGCATGAGCGT 1676  
DB 2691 CAAGCGCGCTTCTTCTTCAACCGGACAGGAGCGGACGAGCGCGCATGAGCGCGGACT 2750  
QY 1677 GTACGATATGATGCTCGCGCTTCCGAGAGCGCTTGAACCTGTGCTGAGAGCTTCA 1736  
DB 2751 GCGCGCTACAGACACCGGTGTGCGGACGCTTGAACAGAGTGTGCGCGCTGAGACG 2810  
QY 1737 GAGGCTGACCGGCGCGCTCGCGAGGATGATGAGGCGGAAACCGGCGACGCTGACGCG 1796  
DB 2811 GACCTTCAACCGGCGCGCTGCGAGAGTCTTGTGCGCGGACGCGCGCGCGCGGCG 2870  
QY 1797 GCTGCTCAGACAGACGCTTCAACCGAGCGCGCGCTTCACTTGAATGCGCTGCG 1856  
DB 2871 CTTGCTGACACAGAGCGCTTCAACCGAGCGCGCGCTTGTGCGCGGAGCTGCGCTCT 2930  
QY 1857 GCGCTGTGCGGTGTGAGAGCGGATGAGAGCGGAGTGTGCGCGCGCATACATGCTGA 1916  
DB 2931 GCGGCTGTGAGAGACTGGGCGCTTGGCGCGGACGATGTGCGGCGCATCTGCTGCG 2990  
QY 1917 GCTGTGCTGCTGCGGTGCGGCGGTGTCTGCTTGAAGACGCGGCTTCTGCTGAGC 1976  
DB 2991 ACTGACCGCGCTTACGCGCGCGGCTTGTGCTGCTGCGCGACGCTGCGCTGCTGCG 3050  
QY 1977 TGCGCGGCGCGCTGATGACGCGCTGCGCGCGCGCGGCGGATGTGTGATGAGCG 2036  
DB 3051 CGCGCGCGCGCGCTGACCGAGGCACTGCGCGGCGCGCGCGCATGCTGCGCTGAGCG 3110  
QY 2037 GCGGAGGCGGATGTGTGCTGCGGTGCGCGCGCGGCGGACGCTGCTGCTGAGTCCG 2096  
DB 3111 GACCGAGGACGAGGTGCGCGCGCAACTGCGGAGCGCGCGCGCGCTGAGCATGCTGCG 3170  
QY 2097 GGTCAAGCTCGCGGACGAGTGTGATGCGGCGCGCGGCGGAAACCGCTGATGCGATGCG 2156  
DB 3171 GGTCAAGGACCGGAGCGGATGTGTCTTCCGCGGAGGAGCGCGCTGCTGAGACTGCG 3230  
QY 2157 GCGCGGATGAGCGCGCGCGCGGCGGAGCAACGAGCGCTTCACTGCTTCACTGCTTCA 2216  
DB 3231 GCGGAGTGTGCGCGCGCGCGCGGAGACGAGGAGCTGCGGCTGACCGCTTCACTGCT 3290  
QY 2217 CTACCGCTTCACTGCGCGCGGATGCTGAGAGCGCTTGGGCGGTGTGCGGAGTGTGAGCT 2276  
DB 3291 CTGCGCTCACTTGAAGCGCATGACGAGGCGTGTGCGCGAGGTGTGACAGAGGAGTCTTCA 3350  
QY 2277 CCGGCGCGCGTGTGATGCTGCTGTGCTGAGCAATCTGAGCGGAGGCTTGCACAGAGAGT 2336

Db 3351 CAGCGCGCGTCCCTCCCGGTGCTCCACGCTCACCGGGCCCCCTCACGACGACT 3410  
Qy 2237 GAGCTCGCGGGCTATTGGTGGCGCCAGCGCGAAGAGTGGTGGCTTGGGAGTAGT 2396  
Db 3411 CCGGAGCGGGAACCTGGGTGGCGGACGCTCCGGGACGGGCTTCCACGACGCGT 3470  
Qy 2397 GAAAGCGCTGACGCGCGGGGTGGGAGCACTTCTGAGGTGGTCCGAAATCGACGCT 2456  
Db 3471 CCGGGCCCTGGCGACCGCGGGGCCACGCGTCTTGAGGTGGGCCCCGGGGGTGCT 3530  
Qy 2457 GCTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2513  
Db 3511 GACGCGCCCGGACCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3590  
Qy 2514 GCGCGCTGGGGGTGACGAGCGCGGACCGGTGCTGAGGCGCTGCGGGGCTTGGGCGCT 2573  
Db 3591 GCGGCGCGCGACGCGCGGACCGGATCGGTGCTGACGCGCGCTGCGGACCGCACGAT 3650  
Qy 2574 CCGTGGCTGTGCTCTCTGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2633  
Db 3651 CCGGCTCTCTGCGCGGTGGGACCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3710  
Qy 2634 CACGTAACCTTGGGACGCGCGGACGCTTACTGATGACACGAAAGCCG 2680  
Db 3711 CACGTACGCTTCCACGCGCGGCACTACTGCTGCTGCGGGCATGGCG 3757

## RESULT 7

US-10-267-255-96  
Sequence 96, Application US/10267255  
Publication No. US20030124689A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D  
APPLICANT: Mao, Y  
APPLICANT: Varoglu, M  
APPLICANT: He, M  
APPLICANT: Sheldon, P  
TITLE OF INVENTION: Mltomycin biosynthetic gene cluster  
FILE REFERENCE: 600.456U31  
CURRENT APPLICATION NUMBER: US/10/267,255  
PRIOR FILING DATE: 2002-10-09  
PRIOR APPLICATION NUMBER: US 09/266,965  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: US 08/624,447  
PRIOR FILING DATE: 1996-08-19  
PRIOR APPLICATION NUMBER: PCT/US94/11279  
PRIOR FILING DATE: 1994-10-06  
PRIOR APPLICATION NUMBER: US 08/133,963  
PRIOR FILING DATE: 1993-10-07  
NUMBER OF SEQ ID NOS: 145  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 96  
LENGTH: 18331  
TYPE: DNA  
ORGANISM: Streptomyces lavendulae  
US-10-267-255-96

Query Match 19.1%; Score 814.2; DB 14; Length 18331;  
Best Local Similarity 57.6%; Pred. No. 6; 8e-170;  
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;  
Qy 18 CGAGGCGCAGCGCAAGATCGATGCGATGCGGAGGAGTGGCGTGGCGCGGTGG 77  
Db 1077 CGAGGACGCGGCGCGGAGCCCGTGCAGATCGTGGGCAATGGCGTCCCGTCCGCGGGA 1136  
Qy 78 CGTATCATCTGACGCGGCTTCTGACGCTCTCTCGAGGAGCTGCGCGACCGTGGGCG 137  
Db 1137 CGTGCATCTGCGCGGACGACCTGTGCACTGTGTGCGCGAGGCGCGGAGCGCGTACCGA 1196  
Qy 138 AGTCCCGCGGAC--GCTGGAGTACAGCAGCGTGGTTGATCCGACCCCGATGCCCT 194  
Db 1197 GTTCCCGCGGACCGGGGCTGGGACGTGACGCGCTTACGACCCCGAGCGGGGACCC 1256

Qy 195 GGGAGAGAGCCCGTTACGCGCGCATCTTCTTGAGGACGTAACCTGCTTGAAGCCTTC 254  
Db 1257 GGGAGAGAGCATACCGCGGCGCACCGCGGCTTCTTAAGAGCGCCCGGATTCGAGCGCCG 1316  
Qy 255 CTCTTGGCATCTCTGCTGCGGAGCGCTGCGGATGAGCCTTGACATCTGATCTTCT 314  
Db 1217 CTCTTGGCATCTCTGCTGCGGAGCGCTGCGGATGAGCCTTGACATCTGATCTTCT 1376  
Qy 315 GGAAGTGTGCTGGGAGGCGCTGAGGAAAGCGCGGATGAGCTCATGCGGCTGCTGCTGAC 374  
Db 1377 GGAAGTGTGCTGGGAGGCGCTGAGGAAAGCGCGGATGAGCTCATGCGGCTGCTGCTGAC 1436  
Qy 375 GAAAGCGGAGTGTTCATGCGGATGCGGCGCTGCGGATGAGGCGCGCTGCGGAGC 434  
Db 1437 GGAAGTGTGCTGGGAGGCGCTGAGGAAAGCGCGGATGAGCTCATGCGGCTGCTGCTGAC 1496  
Qy 435 GACGCGCTGCGGAGAGTCAAGCTCATGCGGCTGAGGAGCATGCGGAGCTGCGGAGC 494  
Db 1497 GCGGAGCGTCCGAGGAGCTTC-----ATCGGAGCGGCAACTCGCGACATCTCTC 1550  
Qy 495 GGGCGGATCTGATGCGCTCGGAGTGGGAGGCGCGGCTGCGGAGTGAATGAGCCTTA 554  
Db 1551 CCGCGGCTGCTTACCTTACCTTGGCTTGAAGGCGCGGCTGCTGCTGCTGCTGCTGCT 1610  
Qy 555 TTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 614  
Db 1611 CTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1670  
Qy 615 CACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674  
Db 1671 CCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1730  
Qy 675 GAAAGCGGCGGCTGCGCGGAGCGGATGCTGCAAGCATTTCTGCGGAGGCGGATG 734  
Db 1731 CCGCGAGGCGGCTGCGGCGGCGGAGCGGCTGCAAGCTTCTGCGGAGGCGGAGG 1790  
Qy 735 GTTGCAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 794  
Db 1791 CACACCTGCTGCTGAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1850  
Qy 795 GAAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 854  
Db 1851 CCGCGGCTGAGCGGCTGAGCGGCTGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1910  
Qy 855 CAGGCTTCTGACCGTGTGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 914  
Db 1911 CCGGCGCTGACCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1970  
Qy 915 GAGCGAGGCTGCGCGCGCTTCTGAGGAGTATGTCAGAGGACAGCGGACAGCGGAGC 974  
Db 1971 CAAAGCAGGCTGACGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2030  
Qy 975 GCTTGTGACCCCATTCGAATTCGAAGCTTGAATGCGGTATACGAGCTCGGAGGAGT 1034  
Db 2031 GCTGAGGAGCCCATTCGAATTCGAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2090  
Qy 1035 CGCAAGCGGCTGCTGATTCGGGTGGTGAAGACCAACTTGGGAGCTCTGAGTATGCTC 1094  
Db 2091 GGGAGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2150  
Qy 1095 GGGAGTCACTGGGCTGCTGAGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1154  
Db 2151 CCGCGCGGCGGCGGCTCATCAAGATGATGATGATGATGATGATGATGATGATGATGAT 2210  
Qy 1155 CCTCAGCGCGAGCGCTGAACCCCGGATCTCATGAGGATCTTCTGAGTACCGCTAC 1214  
Db 2211 GCTGACCTTCAAGAGCCACCCCGGCTGATGATGATGATGATGATGATGATGATGATGAT 2270  
Qy 1215 GCGGCGCGGAGCACGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1274  
Db 2271 GACCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2330





441 AGTGGGTCGCGCCACCAAGGGGTACGGGTGACCGCGGTGACGGTGCAGGGGCGCGG 520  
449 AGATGACGCTCATGCGGGGCTGAGGAGCAATGCCAGCGTCGAGCGGGCCGAATCTCGT 508  
521 AG-----GGGTCCGGGCTGACCGGAGCGCGGACGCGGTCTTGTCGGAAGCAATCTCGT 574  
509 ATGCCCTCGGGCTGCGAGGAGCGGTGTGTCGCGGTGAATACCGGCTTAATGTCCTCGTGG 568  
575 ACCTGCTCGGGCTGAGGGATCCGGCCCTGACCGTGCAGACCGGCTGCTGTCTCGTGG 634  
569 TGGCGGTCATGTGGCTCTGTCAAGACTTGCCTCCGGGGAATGCTTCACGGGCTTGGCTG 628  
635 TGGCGGTACACTTCGCGGTGACGGGCTGCGCGGGGCGAGTGCAGGCTGCGACTGGCGG 694  
629 GTGGGGTATCGCTAATGTTGTGCGCGAGAACCTTGCTGTGGCTCTGAAGACCGGGGCGG 688  
655 GCGGGGTGCGCTTAATGCCGACCGGCGGACATTGTAAGTTCTCCGGGACGCGGGGCG 754  
689 TGGCGAGGAGCGGTGCTGCAAGGCAATTTCCGCGGAGCGCGATGGGTTCCGACGAGCG 748  
755 TCGCGGCGGAGCGGGGCTGCGCGGGGCTTCGGGGCGGGGCGGACCGGACCGGCTGGCGG 814  
749 AAGGTCGCGCTCGTGTCTCTCAAGCGGCTCAATGAGACCGCGCGGACCGCGATCGGA 808  
815 AAGGTCGCTGTGTGCTGTCTGCAACGAGCTCTCCGACGCGGTGCGGACCGGCTGGG 874  
809 TATTGGCGGTGATTCGAGGATTCGCGATCAATCAAGAGGTGCGAGCGCGGTGACGCG 868  
875 TGTGAGGCTGATTCGAGGATTCGCGGATTCGCGATCAAGAGGAGCGGAGCGAGCGGCTG 934  
869 TGGCGAGCGGAGCTGCCAAGAAATGTCGTGAACCGGGCCCTGCGGAGCGAGCGGCTCG 928  
935 CCCGAGGCGGCGCGCGGAGCGGAGCGGATTCGCGGAGCGGCTGACCGACCGCGGCTCG 994  
929 CCGGCTCTTCGCTGAGGATTCGAGGAGCAACGCGGACCGGAGCGAGCGCTTGTGACCG 988  
995 GCGCGGACCAATTCACGCGGTGCGAGGCGGACGCGGACCGGAGCGGAGCGGAGCGG 1054  
989 TCGAAATCAAGCTGATGAGTGCATGAGTACGCGCTGCGGAGGAGTTCGCGACCGCGGCTG 1048  
1055 TCGAGGCGGAGCGGCTGATTCGCGGCTGAGCGGAGCGGAGCGGAGCGGAGCGGCTGCT 1114  
1049 TGATCGGCTGAGTGAAGCAACAATTGAGCACTGATGATGCTGCGGAGTCACTGAGG 1108  
1115 GGGCTGCGCTGATGAAGTGAACATCGGAGCAAGCCGAGCGGCGGCGGCTGCGGCGG 1174  
1109 TGCTGAAGGTGCTTGTTCCTTCAAGCAAGGAGCAATTCCTGCGCACTTCCACGCGCAG 1168  
1175 TGATCAAGATGCTCTGCGGATGCGGAGCGGAGCGCTCCACCGCTGCGACGCGCAG 1234  
1169 CGCTGAACCCCGGATTCATAGGGGTGATCTTCGCGTACCGTCAAGCGGCGCGCGGAC 1228  
1235 TCCCGACCCCGTGTGATGATGCTGCGCGGTGCTGCGGCTGTCACCGGAGTGTGTC 1294  
1229 CGTGGCGGAGCTGG---AATAAGCGCGAGCGGAGCGGAGTGAAGTCTGTTCCGATGAG 1285  
1295 CTTGGCGGCGGTGCGCGGAGCGGCGGCGGAGGATCTTCGCGCTTCGCGGAGTGAAG 1354  
1286 GAGCAACGCGGAGCTGCTGAGAGAGCGCGCGGCGGAGCTGCAACCGCGGAGCG 1345  
1355 GACCAACGCGGAGCTGATGCTGAGAGAGCGCGCGGCTGCGGATCCGAGTCCGAGC 1414  
1346 C-----GAGGAGCGGAGAGCTGCTGCTGCTGCGGAGAGAGCGGCTGAGCG 1396  
1415 CACCCAGAGGAGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1474  
1397 TGAATGCAAGAGCGGCGGCTGCGGAGCAATCTGAGAGCTTAACCTTTCGAGTGTG 1456  
1475 TGGCGGCGGAGCGGAGCGGCTGCTGAGCACTTGGCGGAGCGGAGCGGAGCGGAG 1534  
1457 GCGATGCGGCTTCACTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1516

1535 TGAAGTGGATTCCTCCGCGACCAAGCGGCGCGGCTGAGAGCAAGCGGCGGCTGTCG 1594  
1517 CGGCGAGCTTCAGAGGAGGAGGCTGCGGAGCGGCTGAGAGCTGCGGCGGAGGAGAGAG 1576  
1595 TGGCGGCGGAGCGGAGCGGCTGCTGCGGCTGCTGCTGAGAGCGGCTGCGGCGGAGCG 1654  
1577 GCGCGGCTGCGGAGGAGTTCGCGGATTCCTTCAAGCGGAGGAGGAGGCTGCTTCTTCA 1636  
1655 CGGCGGCGGCTGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1711  
1637 CGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1696  
1712 CGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1771  
1697 TCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1756  
1772 TCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1831  
1757 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1816  
1832 TCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1891  
1817 TCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1876  
1892 ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1951  
1877 GTTGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1936  
1952 GGGTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2011  
1937 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1996  
2012 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2071  
1997 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2056  
2072 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2131  
2057 CTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2116  
2132 CCGTCTGCG-----ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2185  
2117 TGATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2176  
2186 TGTGATGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2245  
2177 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2236  
2246 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2305  
2237 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2296  
2306 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2365  
2297 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2353  
2366 TGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2425  
2354 GGGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2413  
2426 GGGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2485  
2414 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2473  
2486 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2545  
2474 CTTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2530  
2546 AGTGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2605  
2531 AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2590  
2606 AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2665





Db 44178 GCCCTGACCGACCTGCGGGGCTCTCCGCGGGCTGTGTGCTGACCCCGGGGGGGGCTGCC 44237  
Qy 3127 TTTGCTACCCACGTACAACAGTCGCGTGGCTGTGCTGCTCGGCTTCAGGCGCTTCG 3186  
Db 44238 GTGAGCGCCCGGAGAGTCCCGTGTGCGCCGAGCCAGTTGTGGGGCTTCGGCGCGTG 44297  
Qy 3187 GGGATCGAGCGGCGCCCATGCGCGGTGCTGACCGGCAATGTAGCGCGCTTCGACAGA 3246  
Db 44298 GCGCGGCTCGAATCTCCCGGTCTGTGGGCGGCGCTGTGTGACCTTGCGGTGACCCGAGC 44357  
Qy 3247 ATAGCCCGCTTCAGCGCGGGAGCGGGTGTGATCATGCGGCGAGCCGCGGGTGTGCT 3306  
Db 44358 GACCGGAGCTGGGACCTGTGTGGGGCGCGCTGCGCGCGCCGAGGACAGAGTCCGCGT 44417  
Qy 3307 CTCGCCCGGTGTGATGGGCGAGCAGTGGGAGCCGAGGTCCATGCGAGCGCGCAGC 3366  
Db 44418 CGGGGGCGGTGTGAGCCCGCGCGCTGTGCTCCGCGCGCGCGCGCGCGAGCG 44476  
Qy 3367 CCCGAGAACCGGCGCTGAGTGTGGGGGTGTGGGTATGTAGCGGATTCGCCGCTCG 3426  
Db 44477 GGCCTGTGCTCGCGCGGACAGTCTGTGTGACCGCGGTACGCGCGCTCGCGCGCA 44536  
Qy 3427 GACCGGTTGTGCGCGGACGTGC 3448  
Db 44537 CACGGCGCGCTGCGCGCGCGC 44558

RESULT 10  
US-10-212-962-2  
; Sequence 2, Application US/10212962  
; Publication No. US20030114450A1  
; GENERAL INFORMATION:  
; APPLICANT: Kosan Biosciences, Inc.  
; APPLICANT: Santil, Daniel  
; APPLICANT: Myles, David  
; APPLICANT: Zq, Tian  
; APPLICANT: Hutchinson, Richard  
; APPLICANT: Johnson, Robert  
; APPLICANT: Zhou, Yi-Qing  
; APPLICANT: Feng, Li  
; TITLE OF INVENTION: BENZOQUINONE ANZAMYCINS  
; FILE REFERENCE: 30062-20075.00  
; CURRENT APPLICATION NUMBER: US/10/212,962  
; PRIOR FILING DATE: 2002-08-05  
; PRIOR APPLICATION NUMBER: US 60/310,779  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/389,255  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/393,929  
; PRIOR FILING DATE: 2002-07-03  
; PRIOR APPLICATION NUMBER: No. US20030114450A1 yet assigned  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 6360  
; TYPE: DNA  
; ORGANISM: Streptomyces geldanus  
US-10-212-962-2

Query Match 18.0%; Score 765.8; DB 14; Length 6360;  
Best Local Similarity 56.1%; Pred. No. 3.2e-159;  
Matches 1706; Conservative 0; Mismatches 1237; Indels 98; Gaps 10;

Qy 24 CGCAGCCGAGATCCGATTCGATGTCGAGACGAGTTCGCTCCGCTGCGGTGAT 83  
Db 93 CGAGTCCACGAGACCGGTGCGATGTGCGATGTGGGTCCGCGTCCGGGTGAC 152  
Qy 84 CGATTCGACGGGTTTGAAGCTCTCTGAGGAGCTGCGCGACACCGTGTGGCGAGTCCC 143  
Db 153 CGACCCGAGGTGCTGTGTGAGAGTGTGTGACGAGGGCGGAGCGCGATCGGCGCTTCCC 212  
Qy 144 CGCCGAGAC--GCTGGATGCAAGCGGTGTTATCCGACCCCGATGCCCGGGGAGA 200

Db 213 CACGATCGGAGTGGAGCTTGAGAACCTCTTGTGATCTCCATCCGAGCGCTGGGCA 272  
Qy 201 GACGCCCTTACGCGCGCATCTTCTGTAGCGAGCATGACCTGTTCGAGCGCTCTTCT 260  
Db 273 GTCTACGTTAGCGAGGCGGGGTTCTTGAGGGGGCGGGATTCGAGCGCGCTTCTT 332  
Qy 261 CGGCATCTTCGCTCGCGAGGCGGTGCGGATGAGACCTTGACATGATGATCTTGTGAGAT 320  
Db 333 CGGATCTTCGCGCGCGAGGCGCTGTGCTGACCCGAGAGCGGCTGTCTGTGAGAC 392  
Qy 321 GTGTGAGAGGCGGTGAGAACGCGCGCATGCTTCATCGGCGCTGTGCTGATCGAAGC 380  
Db 393 CGCTGGAGAGCTTCGAGCGGGGAGATGATTCGCGGTGTGTGAGAGGCGCGGACAT 452  
Qy 381 GGGAGTTCATCGGAGTCCGCGCGCTCCGAATATAGGCGCGCTGCGCGAGCGAGCGC 440  
Db 453 CGCGGTGTTCCGCGGGGCGAGCGGCGAGGGGTACGCGCGGCTCGGGTGA-----GCG 506  
Qy 441 GTCCGAGAGATCCAGCTCATGAGCGGCTGGGAGCATGCCCGCTGAGAGCGGCGC 500  
Db 507 GCCAAGGCGCTGAGGGGCTATCTGGGGGTGCGGCTTCGCGAGTGTATCTCCGGCGC 566  
Qy 501 AATCTGTATGCGCTCGGGCTGCGAGGCGGTGTGTCGCGGTGATATCGGCTATCTGTC 560  
Db 567 CGTGTGTAACGCTCGGGCTGACCGGTTCGCGGTGACCGTGAACCGCTGCTGCTC 626  
Qy 561 CTCGCTGTGCGCGCTTCTGACCTGTGACAGCTTGGCTCGGGAGATCTCACGCGC 620  
Db 627 CTCGCTGTGCGCGCGCTTCTGACAGCGGTGACAGCGGTGCGCGAGATTCATGTC 686  
Qy 621 CTCGCTGTGCGCGGTATGCTGATGTGTGCGCGAGACCTCTGTGTGCTCTGAGAC 680  
Db 687 GCTGCGCGGTGTGCGCGGTGAGTGGCGCACCGCTCTGTGATGTTCTCCGCA 746  
Qy 681 CGGGCGGTGCGAGGAGCTGCGTGCAGAGCAATTTTCGCGAGGCGGATGAGTTGCG 740  
Db 747 GGTGTGCTGCGCGCGCGAGCGGCGCTGCAATCTTCGCGCGGCGCGCGAGCAC 806  
Qy 741 ACGAGCGAGAGGTGCGCGGTGTGCTCTGACGCGCTGAGAGCGCGCGAGCG 800  
Db 807 CTGTGCGAGAGTGTGCGGCTGTCTGTGAGCGGCTGTGCGAGCGCGCGCGCAAG 866  
Qy 801 CGATCGATATTTGGCGGTGATTCGAGATTCGCGATTCATACAGCGGTGTGAGAGCG 860  
Db 867 CCAGAGTGTGCGGTGATTCGAGATTCGCGAGTCAACAGAGCGCGCTCAAGCG 926  
Qy 861 TCTGACGCTGCGAGAGGAGCTCCCAAGAAATGCTGTGAAAGCGGCTGTGCGAGCG 920  
Db 927 ACTACCGCGCGCAAGCGCGCTCCGAGAGCGGATTCGCGAGCGCTGTCCAGCG 986  
Qy 921 AGCTGCGCGCGCTTCTGTGTGTATGTGTGAGGACAGCGAGCGGCGAGCGCTTGG 980  
Db 987 CGGGTGTGCGGTGCGAGCTGTGAGCGCGGTGAGGCGCGAGCGCGCGCGCTTGG 1046  
Qy 981 TGACCCCATCGAATTCAGCTGTGATGAGCGGTATAGCGGCTGCGGCGAGATGTGCGCAC 1040  
Db 1047 CGACCCCATCGAAGCGCGCGCTTCTGCGACCTACGCGCAAGCGCGCGAGGCGCG 1106  
Qy 1041 GCGGCTGTGATCGGGTGTGAGAGCAACCTTGTGATGATGCTGTGCGGAGAT 1100  
Db 1107 GCGGCTGTGCTGCGCTCTCCATGATTCAGATGCGCACGCGAGCGCGAGGCGGAT 1166  
Qy 1101 CACTGAGTGTGAGAGTGTGTGTCTTTCAGACGCGGAGATTCCTGTGCGACCTTCA 1160  
Db 1167 CGCAGGTATCATTAAGACGCTTATAGCGCACGCGGCTGTGCGAAGACCTTCCA 1226  
Qy 1161 CGCGAGGCGCTGAACCGCGCGATCTCATGAGGTATTTGGCTGACCGCTACGCGGCG 1220  
Db 1227 CGCGAGAGCGCACTTCCAGGTGATGAGTGTGAGCGCGGTGTCTGTGCGCGA 1286  
Qy 1221 CGGACAGGTGTGCGGAGCTGGAATATCGCGCGAGCGGCGGAGTGTGCTGTGCGAT 1280

Db 1287 GAGCGGCGGCTGCGCCGAGACCGGACACGCCCGCGCGCGGATCTCTCTTCGCGCT 1346  
 Qy 1281 GAGCGGCGGACCAACCGCGACGCTGTGTCTGAAAGAGCGCGCGCGGAC----- 1328  
 Db 1347 CAGCGGAGCAACCGCACGCTTCATCTTGAACAGGCGCTTGAAGAACCGAGCGACCGG 1406  
 Qy 1329 -----GTGACACCGCGCGCGCGAGC 1351  
 Db 1407 GAGAACACCGCGATCAGAACCGCGCTACGCTCGCGGAGTCCGCGACCCCGGCGCC 1466  
 Qy 1352 GACCGGCA-----GAGCTCTGTGTGTCTGTCTGCGCAAGACCGCGCTGACGCTTGA 1400  
 Db 1467 GGTGCGCACCGGCGCACGCTGTGTGCGGTGTCTCTCTCGGCGCAATACGAGAGCGCTGCG 1526  
 Qy 1401 TGACAGCGCGCGCGCGCTGCGCGACCATCTGAGACCTAACCTTGTGAGTGTCTGAGCGA 1460  
 Db 1527 TGCCAGGCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1586  
 Qy 1461 TGTGGCGCTTCACTGT 1520  
 Db 1587 CGTGGGCGT 1646  
 Qy 1521 GACCTCGAGGAGGAGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1580  
 Db 1647 CGCGGATGCGCGAGGAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1706  
 Qy 1581 CGGTGCGGT 1640  
 Db 1707 GTCCGCGGT 1760  
 Qy 1641 ACAGGCGCGCGGACGCTGT 1700  
 Db 1761 GACAGGCGAGCGAGCGCGCGGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1820  
 Qy 1701 CGAGCGCTTCA-----CTGT 1748  
 Db 1821 GCGGCGCTTTCACCGCGGT 1880  
 Qy 1749 GCGCGTCCGCGAGGATGT 1808  
 Db 1881 CTCTGTGCGCGAGGATCTTGT 1940  
 Qy 1809 GACAGCGCTTCAACCGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1868  
 Db 1941 GACGAGATTCAACCGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2000  
 Qy 1869 GTCTGT 1928  
 Db 2001 GTCTGT 2060  
 Qy 1929 CTGT 1988  
 Db 2061 CCACTGT 2120  
 Qy 1989 CTTGT 2048  
 Db 2121 GCTGT 2180  
 Qy 2049 TGT 2108  
 Db 2181 GGT 2240  
 Qy 2109 GACACAGGT 2168  
 Db 2241 GGCCTCTGT 2300  
 Qy 2169 GCG 2228  
 Db 2301 GCGGAGGCG 2360  
 Qy 2229 GCG 2288  
 Db 2361 GACGCGGATGT 2420

Qy 2289 GATGCTCTGT 2345  
 Db 2421 GATACCGGT 2480  
 Qy 2346 GCGCTATTGT 2405  
 Db 2481 CAGTACTGT 2540  
 Qy 2406 GACG 2465  
 Db 2541 GCGCGAGAGGCGGT 2600  
 Qy 2466 GGTGCTGT 2525  
 Db 2601 GCGGCGAGAGCTGT 2660  
 Qy 2526 TGACGAGCG 2585  
 Db 2661 CAGAGAGCG 2720  
 Qy 2586 CTCTGT 2642  
 Db 2721 GAGCTGT 2780  
 Qy 2643 TTGTGACG 2702  
 Db 2781 CTTTCAAGACGAGACCTGT 2839  
 Qy 2703 CGGT 2762  
 Db 2840 TGCGGCTGT 2839  
 Qy 2763 CCGGCGGAGCGGT 2818  
 Db 2900 GCGGCGGT 2959  
 Qy 2819 TCGAGCG 2878  
 Db 2960 TCACTCTCGGCGCGGT 3019  
 Qy 2879 ACTGT 2938  
 Db 3020 GCGAGAGACCGGCG 3079  
 Qy 2939 TCGACG 2979  
 Db 3080 CCGGCGACCG 3120

RESULT 11  
 US-10-156-761-2882  
 ; Sequence 2882, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 2882  
 ; LENGTH: 10839

TYPE: DNA  
 ORGANISM: Streptomyces avermitilis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(10839)  
 US-10-156-761-2882

Query Match 17 8%; Score 761.2; DB 14; Length 10839;  
 Best Local Similarity 57.6%; Pred. No. 3,1e-158;  
 Matches 1589; Conservative 0; Mismatches 1103; Indels 66; Gaps 10;

```

QY 26 CAGCCGAAGATCCGATTCGATCGATCGAGCGATGTCGCTGCGCGGATGATG 85
DB 4685 CCGAGGAGACCCGATCGATCGATCGATCGATCGATCGATCGATCGATCGAT 4744
QY 86 ATCTAGCGCGGTTCTGAGCGCTCTGAGAGGCTCGCGGACACCGTGGGAGTCCCG 145
DB 4745 GTCCGAGGACTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4804
QY 146 CCGAAGCGTGGATGACAGAGCGTGTGATCCCGACCCCGATGCCCGGGGAAAGCC 205
DB 4805 CCGACCGAAGCTGCGACCGCGCTGCGCATACGCTTCATCGAGAGTCCGCGGAGCC 4864
QY 206 CCGTACCGCGCGCATCTTCTCTGAGGAGAGTACGCTGCTGCGATCGCTCTTCCGCA 265
DB 4865 G-----GAGGAGTGCCTTCTGCGCGAGGCGCGGACCTTCGACCGCGGCTTCTCGGA 4918
QY 266 TCTGCGCTTCGAGAGCGCTGCGAGATGACCTTCGACATCGACTCTTCTGAGAGTGTCT 325
DB 4919 TCTCCCGCGGTGAGGACACTGCGAGATGATCCGACAGCGGTTGTTGCTGAGAGACTCGT 4978
QY 326 GGGAGCGCTGAGAACCGCGCGATCGCTCCATCGCGCGCTCGTCCGTAACGAAACGGAG 385
DB 4979 GGGAGCGCTGAGAACCGCGCGGATCGCTCCGATCGCTCGCGCGCGCGCGCGCGCGCG 5038
QY 386 TCTTATCGGAGATCGCGCGCTCGCATATGAGCGCGCTGCGCGAGCGAGCGCGCTCG 445
DB 5039 TCTTGTGCGCGGACCGCGCGAGGATGACAGACCGTCTG-----ATGAACTGTCTCG 5092
QY 446 CAGAGATGACGCTCATGCGCGGCTGCGGAGCAATGCCACGCTCGAGCGCGCGCATAT 505
DB 5093 AGGCAAGCGCGCGGATTAACGCTTACCGCGGCGCTGCGGAGCGTGTGTCGCGCGCTCG 5152
QY 506 CGTATGCGCTTGGGCGTGGCGAGGCGCTGTGTCGCGTGTGATGAGCTATGCTCTCTGCG 565
DB 5153 CTTATGCGCTTGGGCGTGGCGAGGCGCGCGCGCTGACGCTGACACCGGTCTCTCTCTGCG 5212
QY 566 TGTGCGCGCTTATGCTGCGCTGCTGAGAGCTTGCCTCGCGGAGATGCTCAACGCGCTCG 625
DB 5213 TGTGCAAGCTGACCTGCGAGGAGGAGGCGCTGCGGCGCGCGAGTCCAGCTGCGCTCG 5272
QY 626 CTGTGCGGATTCGCTGATGTTGTGCGCGACACCTCGTGTGCTCTCGAAGACCCGCG 685
DB 5273 TCGGCGGCTGTGACCTGTCATGCGCACGCGCGGAGCTTCTGTCGAGTTCTGCGCGAGGCG 5332
QY 686 CGCTGCGCGAGGAGCGCTGCGTCAAGGCAATTTTCGCGGAGGCGCGATGCGGTTCCAGAG 745
DB 5333 GTCTGCGGAGAGCGCGCTGCGTCAAGGCAATTTTCGCGGAGGCGCGATGCGGTTCCAGAG 5392
QY 746 GCGAAGGCTGCGCGCTGCTGCTCAAGGCTGCTCAAGGCTGCGGAGCGCGCGCGAGT 805
DB 5393 GCGAAGGCTGCGCGCTGCTGCTGCGCGAGGCTTTCGAGCGCGGAGCGCGAGGCTGCG 5452
QY 806 GGTATTTGCGCGGTATTCAGAGATCCGCGATCAATCAAGAGTGTGCGAGCGCGGTCTGA 865
DB 5453 CGGTCTGCGCGGTGCTGCGGAGTTCGAGCGGTGAACCTCGAGCGGTGTGAGAGCGGTCTGA 5512
QY 866 CCGTCCCGAAGCGGAGCTCCCAAGAAATGTGCTGAAGCGGCGCTGCGGAGCGAGAGCT 925
DB 5513 CCGGCGCGAAGCGGTCTGCGAGAGCGGAGTTCGCGAGCGCGCTGCTGCGCGCGGT 5572
QY 926 GCGCGCGCTTCTGCGGTGCTTATGTCGAGGACACGCGGACGCGGACGAGCTGTGAGACC 985

```

```

DB 5573 TGTGCGCGCGAGCGTGAACATGATGAGAGCGGACAGCGGAGCGTCTGCGCGAGC 5632
QY 986 CCAATGAATCAAGACTGTGAATGCGGTATACGCGCTCGGCGAGATGTCCACAGCCGC 1045
DB 5633 CGATCGAGGACAGCGCGTGTGCTGCGAGCTACGAGAGCGTCCGCGGAGCGGTCCGC 5692
QY 1046 TGTGATCGGAGTGGTGAAGACCAACCTTGCATCTGATGATGAGTGTGCGGAGATCACTG 1105
DB 5693 TGTGCTGAGTTCGATGAGTGAAGTGAACATGAGGACCAACCGATTAACCGCGGTCTCG 5752
QY 1106 GGTCTGAAGTGTCTTGTCTTCTTGAACAGCGGCGAGATTCGCGACCTTCACAGCGC 1165
DB 5753 GGTGATCAAGTCTGATCTGCGCTGCGACAGCGGCGGTGTCCCAAGACCTTCGACGTTG 5812
QY 1166 AGGCGCTAACCCCGGATCTCATAGGAGTATCTTCGCTACCGGTACAGCGCGCGCGGA 1225
DB 5813 ACGAGCCGACCCCGAGGTGAGTGTCTCGGAGGCGCGGTGAGAGTGTGACGAGAGGCG 5872
QY 1226 CACGCTGCGGAGCTGGAATACGCGCGAGCGGCGGAGGTGAGCTGTTCCGCAATGAGCG 1285
DB 5873 GGAAGTGCCTGAGACCGACGAGCGCGCGCGCGCGCGGTGTCTCTGTTCCGAGTCAAGC 5932
QY 1286 GAACCAACGCTGAGTGTCTGAGAAAGCGCGCGCGCGAGTGCACACCGCG----- 1341
DB 5933 GACAGAAACGCGCACTGATTTCTGAGAGCGCGCGCGTGTGAGAGAACGACGCGGTGT 5992
QY 1342 --GCGCGGAGCGACCGGACAGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1399
DB 5993 CTGCGGTGAGCGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6052
QY 1400 ATGACAGCGCGCGCGCTGCGCGACCATCTGAGAACCTTCCGAGTGTGTGCGCG 1459
DB 6053 GGGCGGAGCGCGGACCGCTTGTGCGACCATCTGAGAGAGAGTGTGTGTGTGTGTGTGTGT 6109
QY 1460 ATGTGCGCTTGAATGTGCGACAGACGCGAGCGAGTGAAGACATCGGCTCGCGGTGCG 1519
DB 6110 ACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6169
QY 1520 CGAGCTGAGAGAGGAGGCTGTGCGGACGCTGTGAGACGCTGCGCGCGGAGGACAGAGTGC 1579
DB 6170 GCGCGGATCGGAGAGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6229
QY 1580 CCGGTCGCTGCGAGTATTCGCGATTCCTACCGCGGACATCGCTGCTTCTTCTTCAACG 1639
DB 6230 CCGGATGTGTGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6283
QY 1640 GACAGGAGCGCGACAGCTGCGCATGTGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1699
DB 6284 GTACAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6343
QY 1700 GCGAGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1759
DB 6344 CCGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6403
QY 1760 AGGTATGTGAGCGCAACCGCGACAGCGCGCTGCTGACAGCGAGCTTCA 1819
DB 6404 AGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6448
QY 1820 CCGACGCGCGCTGTTCATCTTGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1879
DB 6449 CCGAGGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6508
QY 1880 TAGAGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1939
DB 6509 TCGCGGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6568
QY 1940 GCGTGTCTTCTGAGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1999
DB 6569 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6628
QY 2000 CCGTCCCGCGCGCGGAGCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2059
DB 6629 CCGTCCCGAGGAGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6677

```



QY	2060	CGGAGGCGCCGACGACGCTGCGGTGCGATGCGCGCGGTACAAGCTCCGACACAGGTCG	2119
Db	6678	-GCTGCCGCTCTTCCCGGGGGGTGTGTGGTGTGCTGCGGTGAACGAGGCCCGCTGTGCTGG	6736
QY	2120	TCATTCGCGGGCGCCGGGCGAACCCCTGTGATGCGATGCGCGGCGGCGATGTCGCGCGGGG	2179
Db	6737	TGCTGTCCGGTGTGATGAGGAGGCCCGGTGACCCGCTCGCCCAAGCGTTTCTGAACGGGCA	6796
QY	2180	CGCGAACCAAGAGCGCTCCACGTTCTTCGATGCGTTTCACTCAACCGCTATGAGCCCGATGC	2239
Db	6797	GGCGGACCAACACAGTTGGCGGCTCAAGTCAAGCGCTTTCACCTCGCGCGGATGAGCGCAATGC	6856
QY	2240	TGAGAGCGCTTGGCGGCTGTGGCCCAAGTGGGTGAGCTACCGCGCGGCGCTCATGCTCTGG	2299
Db	6857	TGGCGGAATTTGGCGGAGACGCTTCGCGCGGTGGGAATTCCTGTCCGCGCGGTATCCCGAGTG	6916
QY	2300	TCAGCAATCTAGCGCGGAAAGGCTTTCACAGACGAGGTAGCTCGCGCGGCTATTGGGTCG	2359
Db	6917	TGTGGAATGTGACCGGTTGGGATGCGCGAGCGCGAGTTTACACACTTCGCGCTTATGGGTGC	6976
QY	2360	GCACACGCGCAGAGAGGTGTGTGCGTTTCGCGATGAGATGAGAGGCGCTGACGCGCGCGTGT	2419
Db	6977	GGCAATGTGCGGGAAGCCGTTGTTTGGCTTTCGCGCATGTGTGTGGCAGAGGTTGTGGGCGGTG	7036
QY	2420	CGGGGACCTTGTGTGAGGTCCGATCCGAATTCGACGCT-----GCTCGGCTGTGCTGT	2473
Db	7037	TGGACAGGTTCTTCGACGCTGGGCTTCGCGCGGGGTGCGCTGACCGCGATGCGCAGGAGACC	7096
QY	2474	CCTCGATGCCGGAAGCGCCCGCGCGCGGTGTGCAATTCGTCCGCGGCTGGGGGTGACGAGC	2533
Db	7097	TCGACCAACACCGGCGCCGACGCGCGGTGCTGCCCTGTACTGTAACCCCGAAGGCCCGAAG	7156
QY	2534	CGGCGACCGTGTCTGAGAGGCGCTTCGGCGGGCTCTGGGCGCGTGGTGTCTTCCTGAG	2593
Db	7157	CGGATGCTGTACCTTCGCGCTTCGGCGGGAATTCACGCGAGGGTGTCCCGGTTCAGTGTGT	7216
QY	2594	CCGGCTTCTTCC--CTCAGGGGGGCGGCGGGTGCCTGCGCCACGTATCCCTTGGCAGC	2650
Db	7217	CCGGGCTGTTCGCGCGGCAACCGGCGCCGCAACCGTGTGACCTGCCACCTAAGCGTTCCAGC	7276
QY	2651	GGGAGCGTACTGTGATTCGACACG--AAAGCCGACGACGCGCGCGGTGTGGCAACGCGGTG	2707
Db	7277	ACAAGCGCTACTGTGCTGTGACTTCGCGCAAGACCGGACGCGGAGCCCGCCGAGATCGGAC	7336
QY	2708	CTCCGGGAGCGGGTTCACGACGAGTTCAGAGAGGGGGGGCGGATGCGCGCGCGACACCG	2765
Db	7337	TGACGCGACCGGAAACCCGCTGTCTGGCGCGCGGATGTGGCGCTGTCCATCTCGACCG	7394

```

1  RESULT 12
2  US-10-156-761-15102
3  Sequence 15102, Application US/10156761
4  Publication No. US20030119018A1
5  GENERAL INFORMATION:
6  APPLICANT: OMURA, SATOSHI
7  APPLICANT: IKEDA, HARUO
8  APPLICANT: ISHIKAWA, JUN
9  APPLICANT: HORIKAWA, HIROSHI
10 APPLICANT: SHIBA, TADAYOSHI
11 APPLICANT: SAKAKI, YOSHIYUKI
12 APPLICANT: HATTORI, YASAHIRA
13 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
14 FILE REFERENCE: 249-262
15 CURRENT APPLICATION NUMBER: US/10/156,761
16 PRIORITY FILING DATE: 2002-05-29
17 PRIORITY APPLICATION NUMBER: JP 2001-204089
18 PRIORITY FILING DATE: 2001-05-30
19 PRIORITY APPLICATION NUMBER: JP 2001-272697
20 PRIORITY FILING DATE: 2001-08-02
21 NUMBER OF SEQ ID NOS: 15109
22 SEQ ID NO 15102
23 LENGTH: 125746

```

```

; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

Query Match      17.8%   Score 761.2;   DB: 14;   Length 125746;
Beet Local Similarity 57.6%   Pred. No. 2.5e-158;
Matches 1589; Conservative 0; Mismatches 1103; Indels 66; Gaps 10;

```

QY	26	AAGCCGAAAGATCCGATTGCGATCGTCGAGAGGAACTTGCCGCTCCGCGGTGGCGTATCG	85
Db	89070	CCGACGACGACCCGATCGTATCGTGGGACATGCGCTCCGCTTCCCGGCGGGTGGCG	89122
QY	86	ATCTGACGGGGTCTGAGACCTCTCTCGAGGGCTTCGCGCACCGCTGGCGAGTCCCG	145
Db	89130	GTCCCGAGGACCTCTGGCGTCTGCTCGTCGACGACCGCGACGAGATGACGGAATTCCCG	8918
QY	146	CCGAAACCTGAGATGACGACGCGTGTTCATCCCGACCCGATGCCCGGGGAAAGACG	205
Db	89190	CCGACCCGAGCGCTGGCACGCGCTGGCCATGACCGCTTCATGAGAGAGTCCGCGGAGCC	8924
QY	206	CCGTTAGCGCGGCACTTTCTCTGAGCGACGATGCGCTCTTCGACGCTCTTTCTTGCGCA	265
Db	89250	G-----GCAAGGTGCTTCTCTGCGCGAGCGCGGACCTTCGACGCGGCTTTCTTGCGGA	89310
QY	266	TCTGCGCTCGCGAAGCGCTCGGATGAGACCCCTGCATCGACTCTTTCCTGAGAGGTGCT	325
Db	89304	TCTGCGCGCGTGAAGCACTGGCGATGATCGCGACGACGGATTCTTGCTGGAGACTTCGT	8936
QY	326	GGGAGGGGCTTGGAGAAACGCGCGCATTCCTTCATCGGCGCTCTGTGGTATCGGAAACGGGAG	385
Db	89364	GGGAGGGGCTTGGAAACGCGCGGGTTTCATCCGCTGGCGCTTCGCGGGACGCGCACCGGCG	8942
QY	386	TGTTTCATCGGAGATCGGCGCGCTCCGCAATATGAGCGCGGCGCTTCGCGCAAGCGACGCGTTCG	445
Db	89424	TCTTCGTCGCGCGGACACCCCGCAGAGTACACGACCGTCTTG-----ATGAATCTGTCG	8947
QY	446	CAGAGATCGACGCTCATGCGGCGGCTGGGAGACGATGCCACGCTCGAGCGGGCCGAATCT	505
Db	89478	AGGCAAGCGCGCGGTTACGCGCTCAACCGGGCGTGGGCAAGGTATGTCGGGCGGGGTCG	89538
QY	506	CGTATGCGCTTCGCGGCTTCGAGGGCGGTGTGTTCGCGGTGATACGGCTTATTCGTCCTTCG	565
Db	89538	CGTATGCGCTTCGCGGCTTCGAGGGCGCGCGCGGTTCAGAGTGCACACGCGTGTCTCTTCGCG	8959
QY	566	TGCGTGGCGCTTTCATCTGGCGCTGTGACAGACTTTCGCGCTCCGGGGAATGCTCCACGCGCTTCG	625
Db	89598	TGCTACGCTGACCTTGACCTTGACGAGGCGAGCGCGCTTCGCGGGCGCGGATGCGACTTCGCGCTTCG	89655
QY	626	CTGGTGGGGATATCGCTGATGTGTCTGCGCGACACCTCTGTGTGGCTCTCGAAGACCCGGG	685
Db	89658	TCGGCGGTTGTAACGCTATGAGCCACCGCGGGAGGCTTCGTGAAGTTCTCGCGGACGAGGGCG	89711
QY	686	CGCTGGCGAGGAATGCTGCTGTCGCAAGGCATTTTCGCGGAGGCGGATGGGTTCCGACGAG	745
Db	89718	GTCCTCGGGGAGACGCGCGGTGTCMAAGGCGTTTCGCGCGGTGTCGACGCGCACGCGGCTTCGCG	89777
QY	746	GCGAAGGCGTGCAGCTGCTGATGCTCTCAAGCGGCTCACTGAGGCCCGCGCGGACGCGCATC	805
Db	89778	GCGAGGGTGTTCGGAATCTGCGCTGCAGCGGCTTTTCGACGCGGTTCGCGGACCGGCGCTC	89833
QY	806	GGATATTTGCGCGTATTTGAGAGATCCGCGATCAATCAACGACGTTGCGACGCGGTCTGA	865
Db	89838	CGGTGCTGGCGGTGTGTCGCGGGTTTCGCGCGGTGAACCTCGGACGGTGTGTCGAAACGCTCTGA	89899
QY	866	CCGTGCGCGAACGCGGAGCTCCCAAGAAATCTGTCTGAAACGCGGCTTCGCGGACGCAAGCT	925
Db	89898	CGGCGCTGAAACGCTCGCTCGCACGACGCGGTGATCTCGGACGCGCTCTGCTTCGCGCGGCT	89955
QY	926	GCGCGCGGTCTTTCGTTGGTATATGTCAGAGCACACGCGCACGCGGACGACGCTTGTGATAC	985
Db	89958	TGTGCGCGCGGACGTCGTGACATGCTGAGAGCGCACGCGGACGCGGGAACCTTCCTCGCGGAC	9001
QY	986	CCATCGAAATTCAGACTCTGAATGCGGTATACGCGCTTCGCGCGAGATGTGCGCACGCGCG	1045

Db 90018 CGATGAGGACACAGGCGTTGCTGCGAGCTACGATCAGGACCGTCCGCGGACCGTCCG 90077  
b|||||  
Qy 1046 TGTGATCGGGTCGGTGAAGAACCACTTGGCATCTGATGATGCGGGGATCATCTG 1105  
Db 90078 TGTGATCGGGTTCGGTGAAGTGAACATCGGCACACCGAGTACGCGCGGGTGTGCGCG 90137  
Qy 1106 GCGTCTGAAGGTGCTTGTGCTTCCCTTCAAGACAGGCAAGTTCTGCGACCTTCCACGCGC 1165  
Db 90138 GTGTATCAAGTCCGTACTGCGCTGCGCAAGCGCTGTGCTCAAGACCTTGCACGTG 90197  
Qy 1166 AGGCGCTGAACCCCGGATCTGATGAGGATGATCTTCCGCTGACCGCTCAAGCGCGCGA 1225  
Db 90198 AGGACCCGACCCCGAGGTGACATGTCGCGCGCGCGCGGTGAGAGCTGCTGACGAGAGGCG 90257  
Qy 1226 CACCGTGCAGGACTGGAATACCGCGCGAGCGGGGGGTGAGCTGTTCCGATGAGCG 1285  
Db 90258 GGGAGTGGCGGAGACCGAGCGGGCGCGCGGGCGGGCGTGTCTTCTGTTCCGGATCAGCG 90317  
Qy 1286 GGAACCAACCGGACGCTGCTGCTGGAAGAGCGCGCGCGCGCGAGCTGACACCGCGCG 1341  
Db 90318 GCACGACCGCGACGCTGATTTGAGAGAGCGCGCTGCTGAGGAAACGAGCGCGGTGT 90377  
Qy 1342 --GCCCGGAGCGACCGGAGAGCTGCTGCTGCTGCTGCGCAAGACCGCGCTGAG 1399  
Db 90378 CTGCGGTTGACGGGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90437  
Qy 1400 ATGCAAGGCGGCGCGCTGCGCGACCATCTGAGAGACCTTCCGCTGAGTCTGAGCG 1459  
Db 90438 GGGCGCACGCGCGCGCTTGGCTGACCATCTGAGGAGAGCGAT--CTGCGCGCGCGCG 90494  
Qy 1460 ATGTGCGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1519  
Db 90495 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90554  
Qy 1520 CGACGCTGAGGAGGAGGCTGCGCGGACCGCTGAGACGCTGCGCGGAGACAGCTGCG 1579  
Db 90555 GCGCGGATCGGAGAGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90614  
Qy 1580 CCGGTGCGGTGCGAGTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1639  
Db 90615 CCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90668  
Qy 1640 GACAGGCGGCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1699  
Db 90669 GTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90728  
Qy 1700 GCGAGGCGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1759  
Db 90729 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90788  
Qy 1760 AGGTATGTGGGCGGACCGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1819  
Db 90789 AGGTATGTGGGCGGACCGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90833  
Qy 1820 CCGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1879  
Db 90834 CCGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90893  
Qy 1880 TAGAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1939  
Db 90894 TCGCGGCGGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90953  
Qy 1940 GCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1999  
Db 90954 GCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91013  
Qy 2000 CCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2059  
Db 91014 CCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91062  
Qy 2060 CCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2119

Db 91063 -GCTGCGGCTGCTTCCCGGGGGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91121  
Qy 2120 TCATGCGGCGCGCGGCAACCCGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2179  
Db 91122 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91181  
Qy 2180 CCGGACCGGAGCGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2239  
Db 91182 GCGGACCGGAGCGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91241  
Qy 2240 TGAAGCGCTTCCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2299  
Db 91242 TGGCGGATTTCCGCGGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91301  
Qy 2300 TCAGCATCTGACCGGAGAGCTTGCACAGACAGAGTGAAGCTTCCGCGGCTTATTTGGGTTC 2359  
Db 91302 TGTGCAATGTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91361  
Qy 2360 GCGACCGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2419  
Db 91362 GCGATGCTGCGGAGGCGCTTGTGCTTCCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91421  
Qy 2420 CCGGACCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2473  
Db 91422 TGAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91481  
Qy 2474 CTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2533  
Db 91482 TCGACACACCGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91541  
Qy 2534 CCGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2593  
Db 91542 CCGACTGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91601  
Qy 2594 CCGGCTTCTTCC--CTGAGGCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2650  
Db 91602 CCGGCTTCTTCCCGGCG 91661  
Qy 2651 GCGAGCGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2707  
Db 91662 ACAAGCGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91721  
Qy 2708 CTTGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2765  
Db 91722 TGACGCGACCGGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 91779

RESULT 13  
US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1.  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:





Qy 1103 CTGGCTGCTGAAGTCTCTTGTCTTCCCTTACAGAGGAGCAATTCTTCGCACTTCACG 1162  
Db 9564 GCGGGGTATCAAGATGCTCCAGAGGCGCTGCGGCAAGCTTGCTGCCGACCCCTCCAGC 9623  
Qy 1163 GCGAGGCGGTAAACCCCGGATCTCATAGGGGTGATCTTTCGCTACCCGTACAGCGCGCC 1222  
Db 9624 CCGAGAACCCAGCCCGCATCGCATGAGGCTCCGCGCGGGTACGGCTGTCTCATCTTCG 9683  
Qy 1223 GGAACACGTGCGCGGACTGGAATACCGCGCAAGGCGGGGGTGAAGTCTGTTCCGATGA 1282  
Db 9684 AGGTGCTTGGCAGAGGAGACCGGCGCGCCCGCGAGACCGGGGTGTCCCTTCGCGCTCG 9743  
Qy 1283 GCGGAGCAACAGCGCGACGTGTGCTGAAGAGCGCGCGCGCGCATGTCACACCGCGCG 1342  
Db 9744 GCGGACCAATGTCCATGTCTGCTCGAAGAGGACCGCGCGCGCGCGCGCGAGACCGG 9803  
Qy 1343 GCGCGAGCGAC-----GCGAGCTGCTGG 1369  
Db 9804 CCGGGAGGCGCGCGCGCTCCCGCGCGCGAGAGGGCGGAGGCGCGCTGCGCTGGG 9863  
Qy 1370 TGCTGTGCGCAAGAGCGCGCTGACGCTGAGATGACAGAGCGCGCGCGCTGCGCGCATC 1429  
Db 9864 TGTCTCCGAGCGAGCGAGCGCGCTGCGGCTCCAGGACCGCGCGCTCCGCGACACC 9923  
Qy 1430 TGGAGACCTTACCTTTCGAGTGTCTGAGCGATGTGTGCGTTCAGTCTGCGACGACGCGCA 1489  
Db 9924 TCTCCCGACCCCGCGCGCGCGCGCGCGCGCATGATGCTTCTCCCTCCCGCGCGCGCG 9983  
Qy 1490 GCGCGATGAGAGCGCGCTCGCGGTGCGCGCGCATGAGAGGAGGCTGCGCGCGACGCC 1549  
Db 9984 CAGCTTTGACACCGCGCGCTGATCGGCTCGGAGCGGGCGCGACCTCCCGCGCGCG 10043  
Qy 1550 TGGAGCGCTGCGCGCGAGGAGAGAGCTGCGCGCGGTGCGGTGCGATTCGCGCATCT 1609  
Db 10044 TGGAGCGCTTGGCGAGAGAGCGAGCGGTCCGCGGTGTGCGGAGTTCGCGCA---CC 10100  
Qy 1610 CAGCGGAGAGCTGCGCTTCTTTCACCGGACAGAGGAGCGAGACGCTGCGCATGAGCGC 1669  
Db 10101 GGGAGCGGAGAGTGGCTTCTTCTTCAACCGGAGGAGGAGGAGCGCGCGCGAGTGGCC 10160  
Qy 1670 GTGGCTGTAGAGATGATGTGTCGCGCTTCGCGAGGCGCTTCGACTGTGCGTGAAGCTGT 1729  
Db 10161 AGGACTTCATGCGCGCCATCTTTCGCGTCCGCGCTCGACAGAGTGAACGACCTTC 10220  
Qy 1730 TCAACAGAGAGCTGAGCGCGCGCTCCGAGGAGTGAATGGGCGCAACCGCGCAGCGTGC 1789  
Db 10221 TCGACCGCTGTGCGCGCGCGCGCTCGCGCGCTGTGAGCGCGCGACCGCGCTGCGCG 10280  
Qy 1790 ACGCGCGCTGCTGACAGAGAGCTTCACCCAGCGCGCGCTGTTCACTTCGAATATG 1849  
Db 10281 AAGCGGCACTCTGAGACCGAGACGAGTACACCGACCGCGCTTTCGCGCTGAGGTG 10340  
Qy 1850 CGCTCGCGCGCTGTGCGGTGTGAGCGGTGTAAGACCGAGTGTGCGCGCGCATGACA 1909  
Db 10341 CGCTCAACCGCTGTGAGAGCACTGGGGAGTGCAGCGCGCTGTGCTGAGGAGCACTCG 10400  
Qy 1910 TCGGTGACTGTGTGCTGCTGCGGTGCGGTGCTTCTCGTTTGAAGACGCGGTTC 1969  
Db 10401 TGGGGAATGAGCGCGCGCGCGCGGTGTGCTGATCTCAACACCGCTGCGCGC 10460  
Qy 1970 TGTGTGCTGCGCGCGCGCTGATGACAGCGCTTCGCGCGCGCGCGCGAGTGTGGA 2029  
Db 10461 TGTGTGCGCGCGCGCGCGAGTGAAGCAAGCGCTTCGCGCGCGCGCGCGAGTGTCCG 10520  
Qy 2030 TCGAGGCGCGAGGCGCATGTGCTGTGCGGTGCGCGCGCGCATGCGAGCGTGTGGA 2089  
Db 10521 TCGCGGCGCGAGAGAGAGGTCCGCGCATGCTGTGCGCGCGCGAGAGACCGCTGCG 10580  
Qy 2090 TCGCGCGGTCAAGCGCTCGAGACAGGTGTGATGCGCGCGCGCGCGAGACCGGTGATG 2149  
Db 10581 TCGCGCGGTGAACGCGCGCGCGGTGTGTATCTCGCGCGCGAGAGAGCGGTGCGCG 10640

Qy 2150 CGATCGCGCGCGCATGAGCGCGCGCGCGCGAGACCAAGCGCTTCACGTTCGATG 2209  
Db 10641 AGCGCGCGCGCGAGCTGCGCGAGAGAGCGCGCGCGCGCGCGCTGCGCGCGCAG 10700  
Qy 2210 GCTTCACTCAACCGCTATAGGCGCGCGAGTGTGAGAGGCGTTCGCGCGCGAGTGG 2269  
Db 10701 CTTTCACTCAACCGCTATAGGCGCGCGAGTGTGAGAGGCGTTCGCGCGCGAGTGG 10760  
Qy 2270 TGAAGTACCGCGCGCGCTGATGCTGCTGAGTCAACAACTGAGCGGGAAG---GTTTGA 2326  
Db 10761 TGGCTTACCGGGAACCGAGGTAGAGTGTGCTTCAAGTCAAGCGCGCGCGCGCGC 10820  
Qy 2327 CAGAGAGTGAAGTCTGCGCGCGCTATTGAGTGCACAGCGCGAGAGGTGTGCTTCG 2386  
Db 10821 CCGGTGAAGTCAACCGCGCGCGCTATGAGTGCAGAGTCCGAGCGCGCTGCGCTTCG 10880  
Qy 2387 CGGATGAGTGAAGCGCTGACCGCGCGCGTGTGGGCACTTGTGAGAGTCCGTCGA 2446  
Db 10881 CCGAGCGGTTCGCGACGAGCAACCGCTCGAGCGCGCACTTGTGAGAGCGCGCGCG 10940  
Qy 2447 AATCGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2506  
Db 10941 ACGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11000  
Qy 2507 CATGCTGCGCGCTGAGCGGTGACAGCGCGCGCGCGCTGCTGAGCGCTGCGCGCGCT 2566  
Db 11001 ---CGGCGATCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCG 11057  
Qy 2567 GGGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2626  
Db 11058 GGGCGCGCGCGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11115  
Qy 2627 CGTGCACGATACCTTGGCGAGCGCGCGCTGCTGATGACAGCAACCGCGAGCG 2686  
Db 11116 CGAGCG 11175  
Qy 2687 CGGCGGTGCGCGCGCGCGCTGCTGCGGAGCGGTGACAGAGGTGAGAGAGGCGCG 2746  
Db 11176 CCGCTACTGCTGCG 11235  
Qy 2747 CGGTGCG 2806  
Db 11236 CTGGAGCGCGGTGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11295  
Qy 2807 GCCGAGAGAGGTGAGCG 2848  
Db 11296 CCACCGGACAGCG 11337

RESULT 15  
US-09-860-846-1  
; Sequence 1, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860.846  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-1  
Query Match 17.4%; Score 742; DB 10; Length 15872;

Best Local Similarity 55.7%; Pred. No. 5e-154;  
Matches 1594; Conservative 0; Mismatches 1215; Indels 53; Gaps 7;

```
QY 26 CAGCCGAGATCGATTGCGATGTCGAGAGCGAGTTGCCCTGCTCCCGGTGCGTATCG 85
Db 8490 CCGAGGACCATCGATCGGATGTTGGGATGCGATGCGCTATCCGGGTGGTGTTCGT 8549
QY 86 ATCTAGAGGGGTTTCGAGCGCTCTCGAGAGGGCTCGGCGACACCGTCCGGGACAGTCCCG 145
Db 8550 CCGCCGAGGACCTGTGGCGGCTGTGTGGCCGAGAGGGACGAGACGATCAGCGAGTTCCCG 8609
QY 146 CCGAAC--GCTGGAGTGCAGCAGCGTGGTTGATCCGACCCCGATGCCCCGGGAGAGA 202
Db 8610 TCAACCGCGGCTGGGACCTGGAGAGCGCTTACGACCCGGATCCGAGTGGAGGGACCA 8669
QY 203 CGCCGGTTACGCGCGCATTTCTCTGAGCGACGTAAGCGCTTTCGACGCTCTCTTCTTTCG 262
Db 8670 CTTACTGCGCGGAGGGCGGGTTCTTGGAAAGCGCGGTGACTTCGACGCGCTTCTTCG 8729
QY 263 GATCTCGGCTGCGGAGGCGTCCGAGTGGACCTTCGACATCGACTTTCGTCGAGAGTCT 322
Db 8730 GATCTCGCGCGCGAGGCGCTGTGTATGACCCGACAGCGGCTGTCTGTGAGAGTCT 8789
QY 323 GCTGGAGGCGCTGGAGAACCGCGCGATCGCTCATCGCGCTCGTCCGTACGAAACCG 382
Db 8790 CTTGGGAGGCGCTGGAGAACCGCGGGGATGACCCGCTCTGCTGCGCGGACACCGCGTGG 8849
QY 383 GAGTGTTCATCGGGATCGGCGCGCTCGGATATGAGGCGCGCTGCCGCAAGCGAGCGCT 442
Db 8850 GTGTCTACTGCGCGCGCGCGCACGCGCTGTACGCTTCGATCCCGGCTGTGTGCCGAG 8909
QY 443 CGGAGAGATGACCGCTCATGCGGGGCTGGGAGACATGCGCACGCTCGGAGGGGCGGA 502
Db 8910 GCTCGAG--GACTATCTGCTGACCGGACGCGCGCGACGCGGATGTCGCGCGCA 8963
QY 503 TCTCTATGCTCCCTCGGCGCTGCGAGGCGCTGTGTGCGCGGTGATACGCGCTATTCGCT 562
Db 8964 TCTCTAGCGCGCTCGGCTCTCGAAGGACCGTCCATGACGCTGGAGAGCGCTCTCTCTCT 9023
QY 563 CGCTGTGCGCTTATCTGCGCTGTGAGAGCTTGGCGCTCGGAGGAGTCTCAACGCGCC 622
Db 9024 CGCTGTGCGCTGATCTGCGCGGTACGCGCGCTGCGGACGAGTGTGCGGCTCGCGCC 9083
QY 623 TGGCGTGGGGGATTCGCGATGTTGTGCGCGAGACCTCGTGTGGCTCTGGAAGCC 682
Db 9084 TGGCGGCGGGGATGCGATGCGGATCGGACGCGCTTCGTGAGGTCTCCGCGACA 9143
QY 683 GGGCGCTGCGCAGGACGCTGCTGCAAGGCAATTTTCGCGGAGGCGGATGGGTTTCGAC 742
Db 9144 AGGGGCTGGCGCGCGACGCGCGCTGCAAGGCGTTCTCGGCGCGCGCGACACCGCGCT 9203
QY 743 GAGCGGAAAGGTGCGCGCTGTGTGTCTCAAGCGGCTCAGTGAGCGCGCGCGAGCGCG 802
Db 9204 GGGCGGAGGGCTCGCGCGCTGTGTGTCTGAGCGGCTGTGAGCGCGCGCGCGCGGCG 9263
QY 803 ATCGGATATTTGGCGGATTCGAGGATCGCGGATCAATGACGCTGGGACACGCGTCT 862
Db 9264 ACAAGCTCTCGGCTGTGACCGGACCGCGGTCAACAGGACGCTGTCTCAACGCGGCT 9323
QY 863 TGAACGTCGCGAGAGGAGCTCCCAAGAAATGCTGCTGAAACGCGGCTCGCGGACGCGAG 922
Db 9324 TGAACGCGCGCGACCGCGCGCGACGCGGATTCGCGAGGCGCTGCGCGAGCGCG 9383
QY 923 GCTGCGCGCGCTTCTGTGTGTATGTGTGAGGACACGCGACGCGGACGACGCTTGTGTG 982
Db 9384 GCTGTCTCCCGGAGGACGCGGCTGTGTGTGAGGCGGACGCGGACGCGGCTTGTGTG 9443
QY 983 ACCCATGGAATTCAGCTCTGGAATGCGGTATACGCGCTCGGCGGAGATGTGCGCACGC 1042
Db 9444 ACCCATGAGGCGCGGCGCGCTGTGTGTGCGCTTCGCGAGGAAACGTTTCGCGGACGAC 9503
QY 1043 CGCTGTGATCGGCTGTGTGAGAACCACTTGGCATCTGTGATGTGCGTGGGAGATCA 1102
```

```
Db 9504 CGCTGTGAGTGGCTCGCTGCGTGAAGTCCAAATTCGCGGATGCCACAGCGCGCGCGTTCG 9563
QY 1103 CTGGGCTGTGAAGTCTGTCTTGTCTTCAGACAGGAGATTTCTGCGACCTTCACG 1162
Db 9564 GCGGCGATCATGAAGTCTTCAGAGGCGTGGCGGACGCTTGTGTGCGCGACCTTCACG 9623
QY 1163 CGCAGGCGCTGAACCCCGGATCTCATGAGGATGATCTTTCGCTGACCGTCAACGCGCGCC 1222
Db 9624 CCGAGAGCCGACCCCGATGCTGACGTGAGTCTCGGCGCGGATGAGGCTGTCTACCTTCG 9683
QY 1223 GGAACCGTGGCGGATCGGAATACGCGCGACGCGGCGGAGGAGCTGTTCGCGATGA 1282
Db 9684 AGGTCCGATGGACGAGACCGGCGCGGCTCGGCGGAGCGGAGTGTCCGCTTCGCGCTCG 9743
QY 1283 GCGGAGCAACGCGGACGTGTGTGTGAGAGGCGCGCGCGGACGTCGACACGCGCG 1342
Db 9744 GCGGACCAATGCGCATGTGTCTTCGAGAGGACCGCGCGCGCGCGCGGAGCGG 9803
QY 1343 CGCGGAGCGAAC-----GGCAGAGCTGTGG 1369
Db 9804 CCGGAGAGCGCGCGCGGCTTCGCGCGCGCGGACGAGAGGCGGAGAGGCTTCGCGCTGG 9863
QY 1370 TCTGTGCGCAAGACCGCGCTGAGCTGATGCAACAGCGGCGCGGCTGCGCGACATC 1429
Db 9864 TGTCTTCGGAACGCGACGAGCGGCGCTGCGGCTTCAGGCGCGGCGGCTTCGCGACAC 9923
QY 1430 TGGAGACTTATCTTTCGAGTGTCTGAGGCTATGTGGCTTCAGTGTGCGACAGACCGCA 1489
Db 9924 TCTCCGACCGCGGCGCGCGCGCGCGCGGATGCAATGCTTCTTCCTGCGCGCAACGCG 9983
QY 1490 GCGCATGAGACACCGGCTCGCGGTCGCGGAGACGTCAGGAGGAGGCTGCGGCGAGCGC 1549
Db 9984 CAGCTTTGACACCGCGCGCGCTGTGATGCGCTGCGGAGCGGAGCGCAATCCGCGCGCGC 10043
QY 1550 TGAAGCTGCGCGGAGGAGACAGAGTCCCGGTGCGGTGCGGAGTCCGATTCCT 1609
Db 10044 TGAAGCGTGTGGCGCAAGACGCGACGCGTCCGCGGTGTGCGCGAGTCCGCGA---CC 10100
QY 1610 CACGCGGAGAGTGGCTTTCTTTTCAACCGGACAGGGGCGGACAGCGCTGGGCAATGGGCG 1669
Db 10101 GGAAGGAGAGATGGCTTTCTTTTCAACCGGAGGAGGAGCAGGCGCGCGGAGTGGCGC 10160
QY 1670 GTGGCTGTAGATATATATGTCGCGGCTCCCGAGAGCGCTTGCAGCTGTGCGTGAAGCTGT 1729
Db 10161 AGGACTGATGCGCGCCATATCTTTCTGCGTCCGCTGACGAGGTGACGACCTTC 10220
QY 1720 TCAACAGAGCTGACCGGCGCGCTCGCGAGGTATGTGAGCGGACGCGCACGCTCG 1789
Db 10221 TGAACCGCTGCTGCGCGCGCGCTCGCGGCTGTGAGCGCGGACCGGCTTCGCGCG 10280
QY 1790 AGCGCGGCTGTGACGACGACGCTTCAACCGCGCGGCTGTTCATCTTTCGAAATATG 1849
Db 10281 AAGCGGCACTCTGAGACGAGACGAGTACACCGGCGGCGCTTTCGCGGTGAGGTGG 10340
QY 1850 CGCTGCGCGCTGTGCGGTGTGTGAGAGCGGATTTGGTCCCGCGCATAGCA 1909
Db 10341 CGCTTCACCGCTGTGAGACATGTGAGGATGCGCGCGGACCTGTGTGGGACACTGG 10400
QY 1910 TCGGTAGCTGTGTGCTGCTGTGCGGCGGCTGTTCCTGTGAGGACGCGGCTTTC 1969
Db 10401 TGGGGAATCTGCGGCGCGCGCACGTCGCGGCTGTGATGCTGACGACGCTTCGCGCG 10460
QY 1970 TGTGTGTGCGCGCGCGGCTGTATGACGCGCTGTGCGCGCGCGCGCGGCGGATGTGTG 2029
Db 10461 TGTGTGTGCGCGCGCGGCTGTATGACGCGCTGTGCGCGCGCGCGCGGCGGATGTGTG 10520
QY 2030 TCGAGGCGCGGAGCGATGTGCTGTGCGGTGCGCGCGCGGACGCGATGTGTG 2089
Db 10521 TCGGAGCGCGGAGGACGAGGTTCGCGGACATCTGTGCGCGCGCGGACGCGCTTCGCG 10580
QY 2090 TCGCGCGGTCAGCGCTTCGAGACGAGTGTATTCGCGGCGCGCGGACCGCTGTGATG 2149
Db 10581 TCGCGCGGTCAGCGCGCGCGGTTCGTGTGTGTATCTCGCGCGGAGGAGCGGTGTGCGG 10640
```

Oy	2150	CGAATCGGCGGCGGATATGCGCGCGCGGCGGCGGAAACAAGGCGCTCAACGTTCTGGCANTG	2209
Db	10641	AGCGCGCGGCGCACTGCCGGAAGAGGCGCGGCACACAGCGGCTTCGCGTTCGCGACG	10700
Oy	2210	CGTTCCATCAACGCGTTCATGAGCCCGCCAGTCTGAGAGCGTTTCGGGCGTGTGCGCAAGTCGG	2269
Db	10701	CGTTCCATCAACCGCTTATGAGCGGCAATGCTCTCCCGGAATTCGCGGAGGTGCGCGCGCGCC	10760
Oy	2270	TGAGCTACCGCGCGCGCGTGCATGCTCTCGGTCAAGCAATGAGACCGGAAAG---GGTTGCA	2326
Db	10761	TGCGCTACCGGAAACCGAGACTGACGCTGCTTTCACAGCTCACAGGAGCGGCGCGCGCC	10820
Oy	2327	CAGACGAGGTGAGCTCCGCGGCGTATTTGGGTGCGCACAGCGCGCAGAGGTGTGCGCTTCG	2386
Db	10821	CCGGTGAATCTACCGGCGCCCGAATCTAGGTGCGCCAGGTCGGTAGCCCGTGGCGCTTCG	10880
Oy	2387	CGGATGGAATGAGAGCGCTGCAACGCGCGCGGTGCGGCACTTTCGTGAGAGTTCGTCCGA	2446
Db	10881	CGGACGCGGTGTCGCAACGACACCGGCTTCGAAACCGGCACCTTTCGTGAGAACCGGCGCG	10940
Oy	2447	AATGACGCTGCTCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2506
Db	10941	ACGCGGTGCTGTGTGGGCAATGACAGAGGAGTCTCGAGAGACGACACCGTGTGCTGCTGCTG	11000
Oy	2507	CATGTGTCGCGCGCTGGGCGTGAACGAGCCGCGCAACCGTGTCTGAGGCGCTTCGCGCGGCTCT	2566
Db	11001	---CGGCGATTCACAAACCGCGGACACCGCGCGGACGATTCGCGGCGCTTCGCGCGCGCTGCG	11057
Oy	2567	GGGCGGTGCGGTGCGCTGTCTCCGTTGGGCGCGGCGCTTCCCGTCAAGGAGGCGCGCGGTGCG	2626
Db	11058	GGGCGGCGCGCGCGCGGTACGCGCGCGGAGCGCGCGCGCGCGCGCGGTGAGCT---GGGCGGATGACGCG	11115
Oy	2627	CGCTGCGCACGTAACCTTGGCAGCGCGAGGCGCTACTGAGATCGACACAGAAAGCGACGACG	2686
Db	11116	CGAGGCGCGCGAGGAGGCGCGCGCGCGCGCTGCAATGCGCCGATCCAGCGCTTCGCGGACCG	11175
Oy	2687	CGGCGCGGTGCGAACCGCGGTGCTTCGCGGAGCGGCTCAACAGAGGTTCAGAGAGGGGCGCG	2746
Db	11176	CGGTTACTGTGCTCGCCCGGAGCGCGCGCGCGGACACCGACGACTGATGTACCGGATTCGG	11235
Oy	2747	CGGTGCGCGCGCGGACCGGCGGACGAGCGCTTGGGCTTCGACATCTCCGCGCGCGGAGCGGAC	2806
Db	11236	CTGGGACCGGCTGCGGCTGTGACCGCGCGGAGCGCGGACCGCGCGCGCTGCGTGGTGTAT	11295
Oy	2807	GCCGCGAAGAGTTCAGAGCGCGCGCGGCGGACCGTTCCTTCGCGCGC	2848
Db	11296	CCACCCCGACGCGCGCTGCGCGGAGCTGTTCGCGGACGCGC	11337

```
Search completed: October 4, 2003, 12:39:42
Job time : 1051.7 secs
```



**THIS PAGE BLANK (USPTO)**



QY 361 GCGCTCTGCTGATGAGAAACGGGAGTGTTCATTCGGGATCCGCCCTCCGAATATGAGGCC 420  
Db 7970 GCGCTCTGCTGATGAGAAACGGGAGTGTTCATTCGGGATCCGCCCTCCGAATATGAGGCC 8029  
QY 421 GCGCTGCGGCAAGCAGCGGCTCCGACAGATTCAGCGCTCATATGCGGGCTGGGACGATG 480  
Db 8030 GCGCTGCGGCAAGCAGCGGCTCCGACAGATTCAGCGCTCATATGCGGGCTGGGACGATG 8089  
QY 481 CCCAGCTGCGAGCGGCGCGAATCTGTATGCTTCGGGCTGCGAGGCGCTGTGTCCG 540  
Db 8090 CCCAGCTGCGAGCGGCGCGAATCTGTATGCTTCGGGCTGCGAGGCGCTGTGTCCG 8149  
QY 541 GTGATATAGGCTATTCCTCTCTGCTGTGTCGCTTCATCTGCGCTGTCAAGCTTGGC 600  
Db 8150 GTGATATAGGCTATTCCTCTCTGCTGTGTCGCTTCATCTGCGCTGTCAAGCTTGGC 8209  
QY 601 TCCGGGGAATGCTTCCAGCGGCTGCTGTGAGGATTCGCTGATGTTGTGCGCGAGCAC 660  
Db 8210 TCCGGGGAATGCTTCCAGCGGCTGCTGTGAGGATTCGCTGATGTTGTGCGCGAGCAC 8269  
QY 661 CTGCTGTGCTCTCGAAGACCGGCGCTGCGCAGAGACGCTGCTGCAAGCATTTTCG 720  
Db 8270 CTGCTGTGCTCTCGAAGACCGGCGCTGCGCAGAGACGCTGCTGCAAGCATTTTCG 8329  
QY 721 GCGAGGCGGATGCTGCTTCGAGCAGAGCGAGGTCGCTGTGTCTTCAAGCGCTC 780  
Db 8330 GCGAGGCGGATGCTGCTTCGAGCAGAGCGAGGTCGCTGTGTCTTCAAGCGCTC 8389  
QY 781 AGTGAAGCCCGCGCGAGCGGCGATTCGATATTTGGCGGTGATTCGAGATTCGCGATCAT 840  
Db 8390 AGTGAAGCCCGCGCGAGCGGCGATTCGATATTTGGCGGTGATTCGAGATTCGCGATCAT 8449  
QY 841 CACGACGCTGCGAGCGGCTGTGACCGGTGCGAAACGGAGCTTCCAAAGAAATCGTGTG 900  
Db 8450 CACGACGCTGCGAGCGGCTGTGACCGGTGCGAAACGGAGCTTCCAAAGAAATCGTGTG 8509  
QY 901 AAAAGGCGCTGCGGAGCAGCGAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
Db 8510 AAAAGGCGCTGCGGAGCAGCGAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8569  
QY 961 GGCACGCGCAGCAGCGCTTGTGTGACCCCATTCGAATTCGAAGCTTCGATATCCGCTATACCGC 1020  
Db 8570 GGCACGCGCAGCAGCGCTTGTGTGACCCCATTCGAATTCGAAGCTTCGATATCCGCTATACCGC 8629  
QY 1021 CTCGGGCGAGATGTGCGCACCGCGCTGCTGATCGGCTCGGTGAGACCAACCTTGCCAT 1080  
Db 8630 CTCGGGCGAGATGTGCGCACCGCGCTGCTGATCGGCTCGGTGAGACCAACCTTGCCAT 8689  
QY 1081 CCTGATATGCTGCGGAGATCATCTGCGGCTGCTGAAGTCGCTTCCCTTCAGACGCGG 1140  
Db 8690 CCTGATATGCTGCGGAGATCATCTGCGGCTGCTGAAGTCGCTTCCCTTCAGACGCGG 8749  
QY 1141 CAGATTCCTGCGCACCTTCACGCGCAGCGGCTGAACCCCGGATTCATTCGAGGATGATTT 1200  
Db 8750 CAGATTCCTGCGCACCTTCACGCGCAGCGGCTGAACCCCGGATTCATTCGAGGATGATTT 8809  
QY 1201 CCGCTGACCTGTCAGCGCGCGCTCGGACACCTGTGCGGACCTGGAATACGCGCGAGCGG 1260  
Db 8810 CCGCTGACCTGTCAGCGCGCGCTCGGACACCTGTGCGGACCTGGAATACGCGCGAGCGG 8869  
QY 1261 GGGGTGAGCTGCTTTCGAGATGAGCGGAGCAACCGCGCAGCTGTGCTGAGAGAGCGCG 1320  
Db 8870 GGGGTGAGCTGCTTTCGAGATGAGCGGAGCAACCGCGCAGCTGTGCTGAGAGAGCGCG 8929  
QY 1321 GCGCGAGAGTGCACACCGCGCGCGCGAGCGAGCGAGAGCTGCTGTGCGCA 1380  
Db 8930 GCGCGAGAGTGCACACCGCGCGCGCGAGCGAGCGAGAGCTGCTGTGCGCA 8989  
QY 1381 AGGACCGGCTGACCTGAGATGACAGCGCGCGCGCTGCGACATCTGAGAGCTTAC 1440  
Db 8990 AGGACCGGCTGACCTGAGATGACAGCGCGCGCGCTGCGACATCTGAGAGCTTAC 9049

QY 1441 CCTTCGAGTGTCTGCGGAGATGTGGCGTTGAGTCTGCGAGCGAGCGCGAGTGGAG 1500  
Db 9050 CCTTCGAGTGTCTGCGGAGATGTGGCGTTGAGTCTGCGAGCGAGCGCGAGTGGAG 9109  
QY 1501 CACCGGCTGCGGAGTGCAGCGCTCGAGGAGGAGGCTGCGGAGCGCTTGAAGCTTGC 1560  
Db 9110 CACCGGCTGCGGAGTGCAGCGCTCGAGGAGGAGGCTGCGGAGCGCTTGAAGCTTGC 9169  
QY 1561 GCGCAGGAGCAGAGTGCCTCGGTCGCTGCGAGTATCCGCCATTCCTACCGGCGAAG 1620  
Db 9170 GCGCAGGAGCAGAGTGCCTCGGTCGCTGCGAGTATCCGCCATTCCTACCGGCGAAG 9229  
QY 1621 CTGCGCTTTCCTTACCGGAGCAGGAGGCGGAGCGCTGAGGATGAGCTGTTCAACAGAG 1680  
Db 9230 CTGCGCTTTCCTTACCGGAGCAGGAGGCGGAGCGCTGAGGATGAGCTGTTCAACAGAG 9289  
QY 1681 GATGTATGTCCTCGGCTTCGCGAGGCGCTTGCACCTGTGCTGAGGCTGTTCAACAGAG 1740  
Db 9290 GATGTATGTCCTCGGCTTCGCGAGGCGCTTGCACCTGTGCTGAGGCTGTTCAACAGAG 9349  
QY 1741 CTGACCGGCGCGCTTCGCGAGGATGTGGCCGGAACCGGCGAGCTGACCGCGCTG 1800  
Db 9350 CTGACCGGCGCGCTTCGCGAGGATGTGGCCGGAACCGGCGAGCTGACCGCGCTG 9409  
QY 1801 CTGACCGAGCAGGCTTCAACCGAGCGCGCTGTTCACTTTCGAATATGCGCTGCGCG 1860  
Db 9410 CTGACCGAGCAGGCTTCAACCGAGCGCGCTGTTCACTTTCGAATATGCGCTGCGCG 9469  
QY 1861 CTGTGCGGCTGTGAGGCTGTAGAGCCGAGATGCTGCGCGGCTTACATTCGATGAGCTG 1920  
Db 9470 CTGTGCGGCTGTGAGGCTGTAGAGCCGAGATGCTGCGCGGCTTACATTCGATGAGCTG 9529  
QY 1921 GTGCTGCTGCTGCTGCGGCGCTGTTCTGCTTGAAGACCGCGGCTTCTGTGTGCTGCG 1980  
Db 9530 GTGCTGCTGCTGCTGCGGCGCTGTTCTGCTTGAAGACCGCGGCTTCTGTGTGCTGCG 9589  
QY 1981 CCGCGGCGCTGATGACAGGCGCTGCGCGCGCGGCGAGATGCTGATGAGCGCGCG 2040  
Db 9590 CCGCGGCGCTGATGACAGGCGCTGCGCGCGCGGCGAGATGCTGATGAGCGCGCG 9649  
QY 2041 GAGGCGGATGCTGCTGCGGTCGCGCGCGCAGCGAGCTGCTGATGAGCTGCGCGCTC 2100  
Db 9650 GAGGCGGATGCTGCTGCGGTCGCGCGCGCAGCGAGCTGCTGATGAGCTGCGCGCTC 9709  
QY 2101 AACGCTTCGAGCAGAGTGTTCATTCGCGGCGCGCGGCAACCGGATGAGGATGCGCG 2160  
Db 9710 AACGCTTCGAGCAGAGTGTTCATTCGCGGCGCGCGGCAACCGGATGAGGATGCGCG 9769  
QY 2161 GCGATGCGCGGCGCGGCGCGGAGCAACAGGCGCTCCAGCTTCGCACTGAGCTCACTCA 2220  
Db 9770 GCGATGCGCGGCGCGGCGCGGAGCAACAGGCGCTCCAGCTTCGCACTGAGCTCACTCA 9829  
QY 2221 CCGCTCATGCGCCCGATGCTGAGAGGCTTCGCGGTCGAGTCCGAGTCCGAGTCCGAG 2280  
Db 9830 CCGCTCATGCGCCCGATGCTGAGAGGCTTCGCGGTCGAGTCCGAGTCCGAGTCCGAG 9889  
QY 2281 CCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340  
Db 9890 CCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9949  
QY 2341 TCGCGGCGCTATTTGGGTCGCGCAGCGCGGAGAGTGTGCTTTCGCGAGTGAAGTGAAG 2400  
Db 9950 TCGCGGCGCTATTTGGGTCGCGCAGCGCGGAGAGTGTGCTTTCGCGAGTGAAGTGAAG 10009  
QY 2401 GCGCTGACCGCGCGGTCGCGGACCTTCTGTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 2460  
Db 10010 GCGCTGACCGCGCGGTCGCGGACCTTCTGTCGAGTTCGAGTTCGAGTTCGAGTTCGAG 10069  
QY 2461 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520  
Db 10070 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10129  
QY 2521 GGGGCTGACGAGCGGCGAGCGGTCGTCGAGGCGCTCGCGGAGCTTGGCGCTGCTGCTG 2580

```

Db      10130 GGGGGTACAGACCGGCGGACCGTCTGAGCGCTCGCGGCGCTCTGAGCGCTGCGTGGC 10189
Qy      2581 CTGGTCTCTGAGGCGGCGCTCTTCCCTCAGGGGGGGGGGGGGCGCGCTGGCCACGTTAC 2640
Db      10190 CTGGTCTCTGAGGCGGCGCTCTTCCCTCAGGGGGGGGGGGGGCGCGCTGGCCACGTTAC 10249
Qy      2641 CTTTGGAGCGCGAGCGCTACTGAGTGCACAGAAACCGACGACGCGCGCGCTGGCGAC 2700
Db      10250 CTTTGGAGCGCGAGCGCTACTGAGTGCACAGAAACCGACGACGCGCGCGCTGGCGAC 10309
Qy      2701 GCGCGTCTCTGAGGCGGCGTCAAGAGAGTGCAGAGAGGGGGGGGGGGGGGGGGG 2760
Db      10310 GCGCGTCTCTGAGGCGGCGTCAAGAGAGTGCAGAGAGGGGGGGGGGGGGGGGGG 10369
Qy      2761 GACCGGCGAGCGCTCGGCTCGACATCCGCGCGCGAGCGGACCGCGGAGAAAGTTC 2820
Db      10370 GACCGGCGAGCGCTCGGCTCGACATCCGCGCGCGAGCGGACCGCGGAGAAAGTTC 10429
Qy      2821 GAGGCGCGCGCGAGCGCTCGGCTCGAGATCGATGACGAGCGGCTTGTATCAC 2880
Db      10430 GAGGCGCGCGCGAGCGCTCGGCTCGAGATCGATGAGCGGCGTGTATCAC 10489
Qy      2881 CTGCGTCTCTGAGGCGGCGTCAAGAGAGTGCAGAGAGGGGGGGGGGGGGGGGGG 2940
Db      10490 CTGCGTCTCTGAGGCGGCGTCAAGAGAGTGCAGAGAGGGGGGGGGGGGGGGGGG 10549
Qy      2941 GACGCGCGGCGGCGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
Db      10550 GACGCGCGGCGGCGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 10609
Qy      3001 CTGCGCGGAGAAAGCCCAACCTCTCGCTCTGCTCGAGAGCGAGTGCAGCGGCGCATGTC 3060
Db      10610 CTGCGCGGAGAAAGCCCAACCTCTCGCTCTGCTCGAGAGCGAGTGCAGCGGCGCATGTC 10669
Qy      3061 GCGGTGGGCGAGGGCGTGAAGCGCTCTGTGGTGGGCGCAACCGGTCATGCGCTTTGCGG 3120
Db      10670 GCGGTGGGCGAGGGCGTGAAGCGCTCTGTGGTGGGCGCAACCGGTCATGCGCTTTGCGG 10729
Qy      3121 GAGGCGGTTGCTACCAAGTCAACCAAGTCAACCAAGTCAACCAAGTCAACCAAGTCAAC 3180
Db      10730 GAGGCGGTTGCTACCAAGTCAACCAAGTCAACCAAGTCAACCAAGTCAACCAAGTCAAC 10789
Qy      3181 CTCTCGCGAGTCAAGAGCGGCGCGCATGCGCTGATGATGATGATGATGATGATGATGAT 3240
Db      10790 CTCTCGCGAGTCAAGAGCGGCGCGCATGCGCTGATGATGATGATGATGATGATGATGAT 10849
Qy      3241 GACAGATTAAGCCCGCTTTCAGCGCGGGGAGCGGGTCTGATCCATGCGCGGCGGCGG 3300
Db      10850 GACAGATTAAGCCCGCTTTCAGCGCGGGGAGCGGGTCTGATCCATGCGCGGCGGCGG 10909
Qy      3301 GTGCGTCTCGCGCGGCGGAGTGGGCGGAGCACTGAGGACCGAGGTCATGCGACGCGC 3360
Db      10910 GTGCGTCTCGCGCGGCGGAGTGGGCGGAGCACTGAGGACCGAGGTCATGCGACGCGC 10969
Qy      3361 GGAAGCGCGGAGAAACCGCGCTACTGAGTGCCTGGGCGTGCAGTATGAGGCGATTCC 3420
Db      10970 GGAAGCGCGGAGAAACCGCGCTACTGAGTGCCTGGGCGTGCAGTATGAGGCGATTCC 11029
Qy      3421 GCGTGGAGCGGTTGCTGCGCGAGCTGCGCGGCGTGAACGCGCGGCGAGGAGTGAAGTTC 3480
Db      11030 GCGTGGAGCGGTTGCTGCGCGAGCTGCGCGGCGTGAACGCGCGGCGAGGAGTGAAGTTC 11089
Qy      3481 GTGCTCAACTCGCTCTGCGGCGAGCTGATGACAAAGTTTCATCTCTGCGATTCGAC 3540
Db      11090 GTGCTCAACTCGCTCTGCGGCGAGCTGATGACAAAGTTTCATCTCTGCGATTCGAC 11149
Qy      3541 GGGCGGTTTGTGAGCTCGGCAAGCGGAGCTGTTACCGGATTAACGAGCTCGGGCTGCGG 3600
Db      11150 GGGCGGTTTGTGAGCTCGGCAAGCGGAGCTGTTACCGGATTAACGAGCTCGGGCTGCGG 11209
Qy      3601 CCGTCTCGGCAATCTCTCTCTCTGCTGCTGATGATCTCGGGGAGATGATGCTGAGCGG 3660

```

```

Db      11210 CCGTCTCGGCAATCTCTCTCTCTGCTGCTGATGATCTCCGGGAGATGATGCTCGAGCGG 11269
Qy      3661 CCGGCGGAGTCCGTCGCTCTTGAAGAGCTCTCGGCTGATTCGCGGAGCGGCTGTTTC 3720
Db      11270 CCGGCGGAGTCCGTCGCTCTTGAAGAGCTCTCGGCTGATTCGCGGAGCGGCTGTTTC 11329
Qy      3721 ACCCTTCCCGCATCGGAGCGCTCCGATGCGCGGTGCGCGAGTGGTTCGAGAGCATG 3780
Db      11330 ACCCTTCCCGCATCGGAGCGCTCCGATGCGCGGTGCGCGAGTGGTTCGAGAGCATG 11389
Qy      3781 GCGGAGCGGAGATCTTGAAGAGCTCTGATCTACAGCTGGGTGATCCCGAGGTCAGATC 3840
Db      11390 GCGGAGCGGAGATCTTGAAGAGCTCTGATCTACAGCTGGGTGATCCCGAGGTCAGATC 11449
Qy      3841 CGTATTCAAACCCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3900
Db      11450 CGTATTCAAACCCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11509
Qy      3901 CTGCGCTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3960
Db      11510 CTGCGCTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11569
Qy      3961 GTCTCGAGGTGCTGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4020
Db      11570 GTCTCGAGGTGCTGCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11629
Qy      4021 CTGCGGATGAGTCTGCTCATGAGCGCGGAGCTGCGGATGATGATGATGATGATGATGATGAT 4080
Db      11630 CTGCGGATGAGTCTGCTCATGAGCGCGGAGCTGCGGATGATGATGATGATGATGATGATGAT 11689
Qy      4081 CTGAGCTGTGAGCGAGCGTTCCTGTCAGCTGCCCAATATGCGCTTGTGGCCAAAC 4140
Db      11690 CTGAGCTGTGAGCGAGCGTTCCTGTCAGCTGCCCAATATGCGCTTGTGTGGCCAAAC 11749
Qy      4141 CTGTGAGTGTCTTCCGACAGCTCTCTCTTGAAGCGGAGTGGCGGAGAACTTACG 4200
Db      11750 CTGTGAGTGTCTTCCGACAGCTCTCTCTTGAAGCGGAGTGGCGGAGAACTTACG 11809
Qy      4201 GCAGGCGTGCMAAACGACTTGTCTCATGCGGCGGAGTCAAGACTGAGAAATCATTTGCC 4260
Db      11810 GCAGGCGTGCMAAACGACTTGTCTCATGCGGCGGAGTCAAGACTGAGAAATCATTTGCC 11869
Qy      4261 CTATGA 4266
Db      11870 CTATGA 11875

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Lignon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelzsch, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
Query Match 100.0%; Score 4266; DB 4; Length 68750;

```

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGCGGATCGTCCCATGAGCGCGACCGAAGATCGATTGCGATGTCGAGCGAGT 60
Db 7610 GTGCGGATCGTCCCATGAGCGCGACCGAAGATCGATTGCGATGTCGAGCGAGT 7669
QY 61 TGGCGTCTGCCCCGGTGGGATGATCGATTGAGCGGGTTCTGAGCGCTCTGAGGGCTCG 120
Db 7670 TGGCGTCTGCCCCGGTGGGATGATCGATTGAGCGGGTTCTGAGCGCTCTGAGGGCTCG 7729
QY 121 CGCGACACCGTCTGGGCGAGTCCCGCGGACGCTGGATGAGAGCGCTGGTTGATCCC 180
Db 7730 CGCGACACCGTCTGGGCGAGTCCCGCGGACGCTGGATGAGAGCGCTGGTTGATCCC 7789
QY 181 GACCCCGATGCCCCGGGGAGAGACGCCGCTTACGCGCGCATCTTCTGAGCGAGTACC 240
Db 7790 GACCCCGATGCCCCGGGGAGAGACGCCGCTTACGCGCGCATCTTCTGAGCGAGTACC 7849
QY 241 TGCCTTCAACGCGCTCTTCTTGGGATCTCGCTCGCGGAGCGCTGCGGATGGAACCTTGA 300
Db 7850 TGCCTTCAACGCGCTCTTCTTGGGATCTCGCTCGCGGAGCGCTGCGGATGGAACCTTGA 7909
QY 301 CATGCACTCTTGTGAGAGTGTGCTGGAGGCGCTGAGAGACGCGCGATGCTCCATCG 360
Db 7910 CATGCACTCTTGTGAGAGTGTGCTGGAGGCGCTGAGAGACGCGCGATGCTCCATCG 7969
QY 361 GCGCTCGTGGATCGAGAAACGAGAGTTCATCGGAGATCGGCGCGCTCGGAATATGAGGCC 420
Db 7970 GCGCTCGTGGATCGAGAAACGAGAGTTCATCGGAGATCGGCGCGCTCGGAATATGAGGCC 8029
QY 421 GCGCTCGCGGAGAGACGCGCGCTCGGAGATTCAGCGCTCATGCGCGGCTGGGAGCATTG 480
Db 8030 GCGCTCGCGGAGAGACGCGCGCTCGGAGATTCAGCGCTCATGCGCGGCTGGGAGCATTG 8089
QY 481 CCCAGCGTCTGGAGGCGGCGGATCTGATGCTCCCTCGGGCTCGAGAGGCGCTGTGTCGCG 540
Db 8090 CCCAGCGTCTGGAGGCGGCGGATCTGATGCTCCCTCGGGCTCGAGAGGCGCTGTGTCGCG 8149
QY 541 GTGATATCGGCTTATCTGCTCGTGGTGGCGCTTCATCTGGCGCTGTCAAGCGCTTGGCG 600
Db 8150 GTGATATCGGCTTATCTGCTCGTGGTGGCGCTTCATCTGGCGCTGTCAAGCGCTTGGCG 8209
QY 601 TCCGGGGAATGCTCAACGCGCGCTGGCTGGGAGATGCGCTGATGCTTCCCGAGCAC 660
Db 8210 TCCGGGGAATGCTCAACGCGCGCTGGCTGGGAGATGCGCTGATGCTTCCCGAGCAC 8269
QY 661 CTCGTGTGCTCTGAGAGACCGGCGCTGGCCAGGAGCGGTGCTGAGAGCAATTTTCG 720
Db 8270 CTCGTGTGCTCTGAGAGACCGGCGCTGGCCAGGAGCGGTGCTGAGAGCAATTTTCG 8329
QY 721 GCGGAGCGGATGAGTTGAGAGAGGCGAGGATGCGCTGTGCTGCTTCAAGCGCTC 780
Db 8330 GCGGAGCGGATGAGTTGAGAGAGGCGAGGATGCGCTGTGCTGCTTCAAGCGCTC 8389
QY 781 AGTGAAGCGCGGAGAGCGGATGAGATTTGGGAGGATTCAGAGATTCGCGATCAT 840
Db 8390 AGTGAAGCGCGGAGAGCGGATGAGATTTGGGAGGATTCAGAGATTCGCGATCAT 8449
QY 841 CACGACGCTGCGAGCGGCTGTGACCGTGCAGACGAGAGCTCCCAAGATGCTGCTG 900
Db 8450 CACGACGCTGCGAGCGGCTGTGACCGTGCAGACGAGAGCTCCCAAGATGCTGCTG 9059
QY 901 AAAGCGGCGCTTGGGAGCGAGGCTGCGCGCGCTTCTGCTGGGTTATGTGAGGACAC 960
Db 8510 AAAGCGGCGCTTGGGAGCGAGGCTGCGCGCGCTTCTGCTGGGTTATGTGAGGACAC 8569
QY 961 GGCACGGGACGAGCGCTTGGAGACCCCATCGAAATCCAGCTCGAAGCGGTATACGGC 1020
Db 8570 GGCACGGGACGAGCGCTTGGAGACCCCATCGAATCCAGCTCGAAGCGGTATACGGC 8629
QY 1021 CTCGGGCGAGATGTCGACGCGCGCTGCTGATTCGGGTCGTGAAGACCAACTTGGCAT 1080
```

```
Db 8630 CTCGGGCGAGATGTCGACGCGCGCTGCTGATCGGCTCGGTAAGACCAACTTGGCAT 8689
QY 1081 CTTGATATGCTGTGGGATCATCTGGGCTGCTGAAGTCTGCTTGTCTCTTCAACACGG 1140
Db 8690 CTTGATATGCTGTGGGATCATCTGGGCTGCTGAAGTCTGCTTGTCTCTTCAACACGG 8749
QY 1141 CAGATTCTGCGACCTTCACGCGGAGGCTGAACCCCGGATCTCATGAGGGGATCTT 1200
Db 8750 CAGATTCTGCGACCTTCACGCGGAGGCTGAACCCCGGATCTCATGAGGGGATCTT 8809
QY 1201 CGGCTGACCGTCAACGCGCGCGCGGACACCGTGGCGGACTGGAATACCGCGCA 1260
Db 8810 CGGCTGACCGTCAACGCGCGCGCGGACACCGTGGCGGACTGGAATACCGCGCA 8869
QY 1261 GGGGTGAGCTGTGCTGCGATGAGCGGAGCAACGCGGACGTGTGCTTGAAGAGCGCG 1320
Db 8870 GGGGTGAGCTGTGCTGCGATGAGCGGAGCAACGCGGACGTGTGCTTGAAGAGCGCG 8929
QY 1321 GCGGCGACGTGACACGCGCGCGCGGAGCGACCGGACAGAGCTGTGCTGTGGCA 1380
Db 8930 GCGGCGACGTGACACGCGCGCGCGGAGCGACCGGACAGAGCTGTGCTGTGGCA 8989
QY 1381 AGGACCGCTCAGGCTCTGATGACACAGCGCGCGCGCTGCGGACCATCTGAGACCTAC 1440
Db 8990 AGGACCGCTCAGGCTCTGATGACACAGCGCGCGCGCTGCGGACCATCTGAGACCTAC 9049
QY 1441 CTTTCCGAGTGTCTGGGCGATGTGCGCTTCAATCTGCGACACGCGGACGCGATGAG 1500
Db 9050 CTTTCCGAGTGTCTGGGCGATGTGCGCTTCAATCTGCGACACGCGGACGCGATGAG 9109
QY 1501 CACCGGCTCGCGGAGGCGGAGCGTGAAGGAGGAGCTGCGGCGACGCTTGAACGCTGCG 1560
Db 9110 CACCGGCTCGCGGAGGCGGAGCGTGAAGGAGGAGCTGCGGCGACGCTTGAACGCTGCG 9169
QY 1561 GCGCAGGAGACAGACGTGCGCGCGCTGCGGACGATTCGCGCATCTTCTCAACGCGGAC 1620
Db 9170 GCGCAGGAGACAGACGTGCGCGCGCTGCGGACGATTCGCGCATCTTCTCAACGCGGAC 9229
QY 1621 CTCGCTTTTCTTTCACCGGACAGGAGGCGGACGCTGGGATGTGGGCTGTGGCTTAC 1680
Db 9230 CTCGCTTTTCTTTCACCGGACAGGAGGCGGACGCTGGGATGTGGGCTGTGGCTTAC 9289
QY 1681 GATGTATGTCGCGGCTTCCGAGAGGCGTTGCACTGTGCGGTGAGGCTGTCAACAGAG 1740
Db 9290 GATGTATGTCGCGGCTTCCGAGAGGCGTTGCACTGTGCGGTGAGGCTGTCAACAGAG 9349
QY 1741 CTCGACCGGCGCTCCGAGAGGATGTGAGCGCGGACCGGACGCTGACGCGCGCTG 1800
Db 9350 CTCGACCGGCGCTCCGAGAGGATGTGAGCGCGGACCGGACGCTGACGCGCGCTG 9409
QY 1801 CTCGACCGAGACGCTTTCACCGGCGGCGCTGTTCACCTTTCGAATATGCGCTGCGCG 1860
Db 9410 CTCGACCGAGACGCTTTCACCGGCGGCGCTGTTCACCTTTCGAATATGCGCTGCGCG 9469
QY 1861 CTGTGCGGCTGTGAGGAGGTGAGAGCGGAGTTGGTCCCGGACCATAGCATCGGTGAGCTG 1920
Db 9470 CTGTGCGGCTGTGAGGAGGTGAGAGCGGAGTTGGTCCCGGACCATAGCATCGGTGAGCTG 9529
QY 1921 GTGGCTGCTGTGCGGCGGCGGTCTGCTTGAAGAGCGCGGTTCTGTGTGCTGCG 1980
Db 9530 GTGGCTGCTGTGCGGCGGCGGTCTGCTTGAAGAGCGCGGTTCTGTGTGCTGCG 9589
QY 1981 GCGGCGGCGCTGATGACGCGGCTGCGCGCGGCGGCGGATGTGTGATTCGAGGCGCG 2040
Db 9590 GCGGCGGCGCTGATGACGCGGCTGCGCGCGGCGGCGGATGTGTGATTCGAGGCGCG 9649
QY 2041 GAGGCGATGAGCTGTGCGGCTGCGGCGGACGACGCTGCGGTGATTCGCGCGCTC 2100
Db 9650 GAGGCGATGAGCTGTGCGGCTGCGGCGGACGACGCTGCGGTGATTCGCGCGCTC 9709
QY 2101 AACGCTCGGACAGGATGATTCGCGGCGCGCGGACCGGATGATTCGCGCGCTC 2160
Db 9710 AACGCTCGGACAGGATGATTCGCGGCGCGCGGACCGGATGATTCGCGCGCTC 9769
```

Qy 2161 GCGATGGCCGCGCGGGGCGCGAACAAGGCGTCTCAAGTCTCGCATGCGTTCCACTCA 2220  
Db 9770 GCGATGGCCGCGCGGGGCGCGAACAAGGCGTCTCAAGTCTCGCATGCGTTCCACTCA 9829  
Qy 2221 CCGCTCATGGCCCGGATGCTGTGAAGCGCTTGGGCGGTGGCCGAGTCCGTGAGTACCGG 2280  
Db 9830 CCGCTCATGGCCCGGATGCTGTGAAGCGCTTGGGCGGTGGCCGAGTCCGTGAGTACCGG 9889  
Qy 2281 CGGCGGTGATGTCCTGCTGAGCAATGAGCGGGAAAGGCTTGACAGAGAGGTAGC 2340  
Db 9890 CGGCGGTGATGTCCTGCTGAGCAATGAGCGGGAAAGGCTTGACAGAGAGGTAGC 9949  
Qy 2341 TCGCGGGCTATTGAGTGGCGGCAAGCGGAGAGGTGTGCTTCCGAGATGAGTGAAG 2400  
Db 9950 TCGCGGGCTATTGAGTGGCGGCAAGCGGAGAGGTGTGCTTCCGAGATGAGTGAAG 10009  
Qy 2401 GCGCTGCAAGCGGCGGCGGTGCGGCACTTCTGAGGTGCTGCGAATGCAAGCTGCTC 2460  
Db 10010 GCGCTGCAAGCGGCGGCGGTGCGGCACTTCTGAGGTGCTGCGAATGCAAGCTGCTC 10069  
Qy 2461 GCGCGGTGCTGCTGCTGAGTGGCGGCAAGCGGCGGCGGTGCTGAGTGGCGGCT 2520  
Db 10070 GCGCGGTGCTGCTGCTGAGTGGCGGCAAGCGGCGGCGGTGCTGAGTGGCGGCT 10129  
Qy 2521 GGGGCTGACAGAGCGGCGAGCGTGTGCTGAGGCGCTGCGGCGGCTTGGGCGGTGAGC 2580  
Db 10130 GGGGCTGACAGAGCGGCGAGCGTGTGCTGAGGCGCTGCGGCGGCTTGGGCGGTGAGC 10189  
Qy 2581 CTGCTCTCTGAGCGCGCTTCTTCCCTGAGGGGGGCGGCGGTGCGCTGCCACGTAC 2640  
Db 10190 CTGCTCTCTGAGCGCGCTTCTTCCCTGAGGGGGGCGGCGGTGCGCTGCCACGTAC 10249  
Qy 2641 CCTTGGGAGGCGGCGCTAGTTCAGACAGAAAGCGAGAGCGGCGGCGGCGGCGGCGAC 2700  
Db 10250 CCTTGGGAGGCGGCGCTAGTTCAGACAGAAAGCGAGAGCGGCGGCGGCGGCGGCGAC 10309  
Qy 2701 CGCGGTGCTCGGAGAGCGGCTGACAGAGAGGTTCAGAGAGGGGCGCGGTGCGGCGGCG 2760  
Db 10310 CGCGGTGCTCGGAGAGCGGCTGACAGAGAGGTTCAGAGAGGGGCGCGGTGCGGCGGCG 10369  
Qy 2761 GACCGGCGAGCGCTCGGCTGACCAATCGCGCGGCGGAGAGCGGCGGAGAGGTC 2820  
Db 10370 GACCGGCGAGCGCTCGGCTGACCAATCGCGCGGCGGAGAGCGGCGGAGAGGTC 10429  
Qy 2821 GAGCGCGCGGCGAGCGGCTCGGCTGACCAATCGCGCGGCGGAGAGCGGCGGAGGTC 2880  
Db 10430 GAGCGCGCGGCGAGCGGCTCGGCTGACCAATCGCGCGGCGGAGAGCGGCGGAGGTC 10489  
Qy 2881 CTGCTGCTTGGGCTGACAGAGAGGCGGCGGCTGAGAGAGGTGAGATGCGCGCTC 2940  
Db 10490 CTGCTGCTTGGGCTGACAGAGAGGCGGCGGCTGAGAGAGGTGAGATGCGCGCTC 10549  
Qy 2941 GAGCGGCGGCGCTGAGCTTCAATGATGTCAGTTCGCGCTGAGGATGTCGCCGAC 3000  
Db 10550 GAGCGGCGGCGCTGAGCTTCAATGATGTCAGTTCGCGCTGAGGATGTCGCCGAC 10609  
Qy 3001 CTGCGGAGAAAGCCCAACCTTCTGCTGCTGAGAGGCGAGTGGCGGCGCATGCTC 3060  
Db 10610 CTGCGGAGAAAGCCCAACCTTCTGCTGCTGAGAGGCGAGTGGCGGCGCATGCTC 10669  
Qy 3061 GCGCGTGGAGAGGCGGTGAGAGGCTGAGTGGGCGCAACCGGATGATGCGCTTCCGCG 3120  
Db 10670 GCGCGTGGAGAGGCGGTGAGAGGCTGAGTGGGCGCAACCGGATGATGCGCTTCCGCG 10729  
Qy 3121 GAGAGCTTGTCTACCAAGTCAACAGTTCGCGTGGCTGTGCTGCTGCGCTTCAAGCG 3180  
Db 10730 GAGAGCTTGTCTACCAAGTCAACAGTTCGCGTGGCTGTGCTGCTGCGCTTCAAGCG 10789  
Qy 3181 CTCTCGGAGATGAGAGCGGCGGCAATGCGGCTGAGTACCTGACGAGATGATGCGGCTC 3240  
Db 10790 CTCTCGGAGATGAGAGCGGCGGCAATGCGGCTGAGTACCTGACGAGATGATGCGGCTC 10849

Qy 3241 GACAGAAATAGCCGCTTTCAGCCGCGGAGCGGCTGATTCATGCGGCGACCGCGCG 3300  
Db 10850 GACAGAAATAGCCGCTTTCAGCCGCGGAGCGGCTGATTCATGCGGCGACCGCGCG 10909  
Qy 3301 GTGCGTCTCGCGGCGGCTGAGTGGCGGCAAGCTTGGAGGCGGAGTTCATGCGGCG 3360  
Db 10910 GTGCGTCTCGCGGCGGCTGAGTGGCGGCAAGCTTGGAGGCGGAGTTCATGCGGCG 10969  
Qy 3361 GGCAGCGCGGAGAAACGCGCTTACCTGAGAGTGTGAGGCGGCTGAGGAGTTC 3420  
Db 10970 GGCAGCGCGGAGAAACGCGCTTACCTGAGAGTGTGAGGCGGCTGAGGAGTTC 11029  
Qy 3421 CGCTCGAGCGGCTGCTGCGGAGTGGCGGCGGCTGAGCGGCGGAGAGTACGCTC 3480  
Db 11030 CGCTCGAGCGGCTGCTGCGGAGTGGCGGCGGCTGAGCGGCGGAGAGTACGCTC 11089  
Qy 3481 GTGCTCAATGCTGCTGCGGCGGAGTATGAGCAAGATTTCATCTCTGCGATTCGAC 3540  
Db 11090 GTGCTCAATGCTGCTGCGGCGGAGTATGAGCAAGATTTCATCTCTGCGATTCGAC 11149  
Qy 3541 GCGCGGTTTGGAGCTCGGCAAGCGACTGTTACGCGGATACAGCTCGGCGGCG 3600  
Db 11150 GCGCGGTTTGGAGCTCGGCAAGCGACTGTTACGCGGATACAGCTCGGCGGCG 11209  
Qy 3601 CCGTTCGCGCAATCTCTCTCTGCTGCTGAGTCTCGGCGGAGTATGCTGAGCGG 3660  
Db 11210 CCGTTCGCGCAATCTCTCTCTGCTGCTGAGTCTCGGCGGAGTATGCTGAGCGG 11269  
Qy 3661 CCGGCGGCGGCTCGGCGGCTTGGAGAGTCTCTCGGCGGATGAGTCTGAGCGG 3720  
Db 11270 CCGGCGGCGGCTCGGCGGCTTGGAGAGTCTCTCGGCGGATGAGTCTGAGCGG 11329  
Qy 3721 ACCCTTCCCCCATGCGAGCGCTCCGATGCGCGGCTGCGGAGTTCGCGAGCATG 3780  
Db 11330 ACCCTTCCCCCATGCGAGCGCTCCGATGCGCGGCTGCGGAGTTCGCGAGCATG 11389  
Qy 3781 GCGGAGCGGAGCATCTTGGAGAGCTGTACTACGCTGAGTGAACCGGAGTCCAGATC 3840  
Db 11390 GCGGAGCGGAGCATCTTGGAGAGCTGTACTACGCTGAGTGAACCGGAGTCCAGATC 11449  
Qy 3841 GGTATTCACACCGAGAGGCGGCGGCTTCCACCGGAGTTCGAGTTCGACAGG 3900  
Db 11450 GGTATTCACACCGAGAGGCGGCGGCTTCCACCGGAGTTCGAGTTCGACAGG 11509  
Qy 3901 CTGCGGTGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3960  
Db 11510 CTGCGGTGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11569  
Qy 3961 GTCTCGAGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4020  
Db 11570 GTCTCGAGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11629  
Qy 4021 CTGCGCATGATCTGCTGATGAGCGGCTGAGTTCGAGTTCGAGGCGGCGGCGGCGG 4080  
Db 11630 CTGCGCATGATCTGCTGATGAGCGGCTGAGTTCGAGTTCGAGGCGGCGGCGGCGG 11689  
Qy 4081 CTGAGGCTGTCAGAGAGTTCCTGTCACGCTCCCAATTCGCTTGGCGGCGGCGG 4140  
Db 11690 CTGAGGCTGTCAGAGAGTTCCTGTCACGCTCCCAATTCGCTTGGCGGCGGCGG 11749  
Qy 4141 CTGTTGATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4200  
Db 11750 CTGTTGATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11809  
Qy 4201 GCGAGGCTGCAAAAGCACTTGTCTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4260  
Db 11810 GCGAGGCTGCAAAAGCACTTGTCTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11869  
Qy 4261 CTATGA 4266  
Db 11870 CTATGA 11875

RESULT 3  
 US-09-567-969-1  
 ; Sequence 1, Application US/09567969  
 ; Patent No. 6355457  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Zirkle, Ross  
 ; APPLICANT: Cyt, Devon  
 ; APPLICANT: Goelach, Joern  
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 ; FILE REFERENCE: 4-30582A  
 ; CURRENT APPLICATION NUMBER: US/09/567,969  
 ; PRIOR FILING DATE: 2000-05-10  
 ; PRIOR FILING DATE: 09/335,409  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 68750  
 ; TYPE: DNA  
 ; ORGANISM: Sorangium cellulosum  
 US-09-567-969-1

Query Match 100.0%; Score 4266; DB 4; Length 68750;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGAGATCGTCCATCGAGCGCCAGCCGAAAGATCCGATGCGATGCGAGCGAGT 60  
 DB GTGGGAGATCGTCCATCGAGCGCCAGCCGAAAGATCCGATGCGATGCGAGCGAGT 7669  
 QY 61 TCCCTGTGCGCGGTGCGGTGATGATGATGAGCGGCTTCTGAGCGCTCTGAGGCGTGC 120  
 DB TCCCTGTGCGCGGTGCGGTGATGATGATGAGCGGCTTCTGAGCGCTCTGAGGCGTGC 7729  
 QY 121 CGCGACACCTGCGCGCGAGTCCCGCCGAGCGCTGGATGAGCGAGCGGTGTTATCCC 180  
 DB CGCGACACCTGCGCGCGAGTCCCGCCGAGCGCTGGATGAGCGAGCGGTGTTATCCC 7789  
 QY 181 GACCCGATGCGCGCGGAGAGCGCCGTTACGCGCGCATCTTCTGAGCGAGTACGC 240  
 DB GACCCGATGCGCGCGGAGAGCGCCGTTACGCGCGCATCTTCTGAGCGAGTACGC 7849  
 QY 241 TCGTTGACGCGCTCTTCTGCGCATCTGCGCTGCGAAGCGCTGCGAGTGAACCTTGA 300  
 DB TCGTTGACGCGCTCTTCTGCGCATCTGCGCTGCGAAGCGCTGCGAGTGAACCTTGA 7909  
 QY 301 CATGCACTCTTCTGAGAGTGTCTGCGAGCGCGCTGAGAGAGCGCGCATGCGCTTCATCG 360  
 DB CATGCACTCTTCTGAGAGTGTCTGCGAGCGCGCTGAGAGAGCGCGCATGCGCTTCATCG 7969  
 QY 361 GCGCTGCTGCGTACGGAAGAGGAGTGTTCATCGGAGATCGCGCTGCGAATATGAGGCC 420  
 DB GCGCTGCTGCGTACGGAAGAGGAGTGTTCATCGGAGATCGCGCTGCGAATATGAGGCC 8029  
 QY 421 GCGCTGCGTACGAGCGCGCTGCGAGAGATGACGCTCATGAGCGCGCTGAGGAGCATG 480  
 DB GCGCTGCGTACGAGCGCGCTGCGAGAGATGACGCTCATGAGCGCGCTGAGGAGCATG 8089  
 QY 481 CCCAGCGTGCAGCGCGCGCGATCTCTGATGCGCTGCGAGAGCGCGCTGAGGAGCATG 540  
 DB CCCAGCGTGCAGCGCGCGCGATCTCTGATGCGCTGCGAGAGCGCGCTGAGGAGCATG 8149  
 QY 541 GTGATACGCGCTTATCTGCTGCTGCGTGCCTGTTACATGCGCTGTCAGAGCTTGGCG 600  
 DB GTGATACGCGCTTATCTGCTGCTGCGTGCCTGTTACATGCGCTGTCAGAGCTTGGCG 8209  
 QY 601 TCCGGGAGATGCTCAACGCGCGCTGCGTGGGAGTATGCGCTGATGTTGTCGCGAGCAC 660  
 DB TCCGGGAGATGCTCAACGCGCGCTGCGTGGGAGTATGCGCTGATGTTGTCGCGAGCAC 8269

QY 661 CTCGATGAGCTCTGAGAGACCCGGGCGCTGCGAGGAGCGGTGCGTCAAGCATTTTCG 720  
 DB CTCGATGAGCTCTGAGAGACCCGGGCGCTGCGAGGAGCGGTGCGTCAAGCATTTTCG 8329  
 QY 721 GCGAGGCGCGATGAGGATTCGAGAGCGAGAGGAGTGCCTGCTGCTCAAGCGCTC 780  
 DB GCGAGGCGCGATGAGGATTCGAGAGCGAGAGGAGTGCCTGCTGCTCAAGCGCTC 8389  
 QY 781 AGTGAAGCCCCCGGAGAGCGCGATGATGATGAGCGGTGATTCAGAGATCCGCGATCAAT 840  
 DB AGTGAAGCCCCCGGAGAGCGCGATGATGATGAGCGGTGATTCAGAGATCCGCGATCAAT 8449  
 QY 841 CACGACGCTGAGAGAGCGGTCTGACCGCTGCGAGAGGAGCTCCAGAAATGCTGCTG 900  
 DB CACGACGCTGAGAGAGCGGTCTGACCGCTGCGAGAGGAGCTCCAGAAATGCTGCTG 8509  
 QY 901 AAACGGGCGCTGAGGAGAGCGAGGCTGCGCGCTGCTGCTGAGGATGATGAGGAGCAC 960  
 DB AAACGGGCGCTGAGGAGAGCGAGGCTGCGCGCTGCTGCTGAGGATGATGAGGAGCAC 8569  
 QY 961 GGCACGCGACGAGCGCTTGTGATCCCATGCAATCCAGCTTGAATGCGGTATACGCG 1020  
 DB GGCACGCGACGAGCGCTTGTGATCCCATGCAATCCAGCTTGAATGCGGTATACGCG 8629  
 QY 1021 CTCGGGCGAGATGTCGCGAGCGCGCTGCTGATGCGGCTGAGAGAGCAACCTTGCCAT 1080  
 DB CTCGGGCGAGATGTCGCGAGCGCGCTGCTGATGCGGCTGAGAGAGCAACCTTGCCAT 8689  
 QY 1081 CCGATGATGCTGCGGAGATCACTGCGCTGAGAGGCTGCTTGTCCCTTCAACACGCG 1140  
 DB CCGATGATGCTGCGGAGATCACTGCGCTGAGAGGCTGCTTGTCCCTTCAACACGCG 8749  
 QY 1141 CAGATTCCTGCGACCTTCAACGCGAGCGCTGAACCCCGGATCTCATGCGGTGATCTT 1200  
 DB CAGATTCCTGCGACCTTCAACGCGAGCGCTGAACCCCGGATCTCATGCGGTGATCTT 8809  
 QY 8750 CAGATTCCTGCGACCTTCAACGCGAGCGCTGAACCCCGGATCTCATGCGGTGATCTT 8809  
 DB CAGATTCCTGCGACCTTCAACGCGAGCGCTGAACCCCGGATCTCATGCGGTGATCTT 8869  
 QY 1201 GCGGTGACCTGACGCGCGCGCGAGCAACCGTGGCGGAGCTGAATACGCGCGAGCGCG 1260  
 DB GCGGTGACCTGACGCGCGCGCGAGCAACCGTGGCGGAGCTGAATACGCGCGAGCGCG 8869  
 QY 1261 GGGGTGACCTGCTTCCGATGAGCGGAGCAACCGCGAGCTGCTGAGAGAGCGCGCG 1320  
 DB GGGGTGACCTGCTTCCGATGAGCGGAGCAACCGCGAGCTGCTGAGAGAGCGCGCG 8929  
 QY 1321 GCGGAGCTGACACAGCGCGCGCGCGAGAGCGAGCGAGAGCTGCTGCTGCTGCGCA 1380  
 DB GCGGAGCTGACACAGCGCGCGCGCGAGAGCGAGCGAGAGCTGCTGCTGCTGCGCA 8989  
 QY 8930 GCGGAGCTGACACAGCGCGCGCGAGAGCGAGCGAGAGCTGCTGCTGCTGCGCA 8989  
 DB GCGGAGCTGACACAGCGCGCGCGAGAGCGAGCGAGAGCTGCTGCTGCTGCGCA 1440  
 QY 1381 AGGACCGGCTGAGCGCTGAGTGAACAGCGCGCGCTGCGAGCATCTGGAAGCTTAC 1440  
 DB AGGACCGGCTGAGCGCTGAGTGAACAGCGCGCGCTGCGAGCATCTGGAAGCTTAC 9049  
 QY 8990 AGGACCGGCTGAGCGCTGAGTGAACAGCGCGCGCTGCGAGCATCTGGAAGCTTAC 9049  
 DB AGGACCGGCTGAGCGCTGAGTGAACAGCGCGCGCTGCGAGCATCTGGAAGCTTAC 9109  
 QY 1441 CTTTCCAGATGCTGAGGAGTGTGAGCTTCACTGCGAGAGCGCGAGCGAGTGAAG 1500  
 DB CTTTCCAGATGCTGAGGAGTGTGAGCTTCACTGCGAGAGCGCGAGCGAGTGAAG 9109  
 QY 9050 CTTTCCAGATGCTGAGGAGTGTGAGCTTCACTGCGAGAGCGCGAGCGAGTGAAG 9109  
 DB CTTTCCAGATGCTGAGGAGTGTGAGCTTCACTGCGAGAGCGCGAGCGAGTGAAG 9169  
 QY 1501 CACCGGCTGCGGTGCGCGAGCGCTGAGAGAGGAGGCTGCGAGAGCGCTGAGAGCTGCG 1560  
 DB CACCGGCTGCGGTGCGCGAGCGCTGAGAGAGGAGGCTGCGAGAGCGCTGAGAGCTGCG 9169  
 QY 9110 CACCGGCTGCGGTGCGCGAGCGCTGAGAGAGGAGGCTGCGAGAGCGCTGAGAGCTGCG 9169  
 DB CACCGGCTGCGGTGCGCGAGCGCTGAGAGAGGAGGCTGCGAGAGCGCTGAGAGCTGCG 9229  
 QY 1561 GCGCAGGAGAGAGAGTGCAGCGCGGTGCGAGTATGCGCGATTCCTCAACGCGAG 1620  
 DB GCGCAGGAGAGAGAGTGCAGCGCGGTGCGAGTATGCGCGATTCCTCAACGCGAG 9229  
 QY 9170 GCGCAGGAGAGAGTGCAGCGCGGTGCGAGTATGCGCGATTCCTCAACGCGAG 9229  
 DB GCGCAGGAGAGAGTGCAGCGCGGTGCGAGTATGCGCGATTCCTCAACGCGAG 9289  
 QY 1621 CTCGCTTCTTCTTCAACGAGAGCGAGAGCGCTGAGGAGCGCTGAGGCTGATC 1680  
 DB CTCGCTTCTTCTTCAACGAGAGCGAGAGCGCTGAGGAGCGCTGAGGCTGATC 9289  
 QY 9230 CTCGCTTCTTCTTCAACGAGAGCGAGAGCGCTGAGGAGCGCTGAGGCTGATC 9289  
 DB CTCGCTTCTTCTTCAACGAGAGCGAGAGCGCTGAGGAGCGCTGAGGCTGATC 9349  
 QY 1681 GATGATGATGCTGCGGTTCGAGAGCGCTTCAACCTGTCGAGAGCTGTTCAACAGAG 1740  
 DB GATGATGATGCTGCGGTTCGAGAGCGCTTCAACCTGTCGAGAGCTGTTCAACAGAG 9349  
 QY 9290 GATGATGATGCTGCGGTTCGAGAGCGCTTCAACCTGTCGAGAGCTGTTCAACAGAG 9349  
 DB GATGATGATGCTGCGGTTCGAGAGCGCTTCAACCTGTCGAGAGCTGTTCAACAGAG 1800  
 QY 1741 CTCGACGCGCGCTTCCGAGAGTATGAGCGAGAGCGAGCGAGCTGAGCGCGCTG 1800



```
Db 9350 CTGACCGGCGCTCCGCGAGGTGATGAGCCGACCGGCGAGCGCTGCGACGCGCGCTG 9409
Qy 1801 CTGACCGAGACAGCTTCAACCCAGCGCGCGCTGTTCACTTCAATATAGCGCTCGCGCG 1860
Db 9410 CTGACCGAGACAGCTTCAACCCAGCGCGCGCTGTTCACTTCAATATAGCGCTCGCGCG 9469
Qy 1861 CTGTGGCGGTCTGTGGGTGTAGAGCCGAGTGTGTGCGCGCCATATAGCATCGGTGAGCTG 1920
Db 9470 CTGTGGCGGTCTGTGGGTGTAGAGCCGAGTGTGTGCGCGCCATATAGCATCGGTGAGCTG 9529
Qy 1921 GTAGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Db 9530 GTAGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9589
Qy 1981 CGGCGGCGCTGATGACGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
Db 9590 CGGCGGCGCTGATGACGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9649
Qy 2041 GAGCGCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2100
Db 9650 GAGCGCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9709
Qy 2101 AAGCGCTCCGAGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
Db 9710 AAGCGCTCCGAGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9769
Qy 2161 GCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2220
Db 9770 GCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9829
Qy 2221 CGCGCTATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2280
Db 9830 CGCGCTATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9889
Qy 2281 CGCGCTATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2340
Db 9890 CGCGCTATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9949
Qy 2341 TGCGCGGCGCTATGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400
Db 9950 TGCGCGGCGCTATGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10009
Qy 2401 GCGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2460
Db 10010 GCGCTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10069
Qy 2461 GCGCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520
Db 10070 GCGCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10129
Qy 2521 GGGCGGTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2580
Db 10130 GGGCGGTGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10189
Qy 2581 CTGGGTCTCTGTGGCGCGCGCTCTTCCCTCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2640
Db 10190 CTGGGTCTCTGTGGCGCGCGCTCTTCCCTCAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 10249
Qy 2641 CTTTGGACGCGAGCGCTATCTGATGTGACACAAAGCCGACGACGCGCGCGCGCGCGCGCGAC 2700
Db 10250 CTTTGGACGCGAGCGCTATCTGATGTGACACAAAGCCGACGACGCGCGCGCGCGCGCGAC 10309
Qy 2701 CGCGGTGCTCCGGGAGCGGGTCAAGACAGTGTGAGGAGGGGGGGGGGGGGGGGGGGGGGGGG 2760
Db 10310 CGCGGTGCTCCGGGAGCGGGTCAAGACAGTGTGAGGAGGGGGGGGGGGGGGGGGGGGGGGGG 10369
Qy 2761 GACCGCGAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2820
Db 10370 GACCGCGAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10429
Qy 2821 GAGCGCGCGCGCGAGCGGTCCGCTTCCGGCTGAGATGTGATGAGCGAGCGGTGCTTGTATCAC 2880
Db 10430 GAGCGCGCGCGCGAGCGGTCCGCTTCCGGCTGAGATGTGATGAGCGAGCGGTGCTTGTATCAC 10489
Qy 2881 CTGCTCTTCTGAGGTCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2940
Db 10490 CTGCTCTTCTGAGGTCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10549
Qy 2941 GACCGCGCGCGCGCTGAGCTTCAATGATGTTCAGCTGTGCGCTGTGGGACATGTGTCCGACAC 3000
Db 10550 GACCGCGCGCGCGCTGAGCTTCAATGATGTTCAGCTGTGCGCTGTGGGACATGTGTCCGACAC 10609
Qy 3001 CTGCGCGGAAAGCCCAACCTCCGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3060
Db 10610 CTGCGCGGAAAGCCCAACCTCCGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10669
Qy 3061 GCGGTGAGGAGGAGCGTGAACGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3120
Db 10670 GCGGTGAGGAGGAGCGTGAACGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10729
Qy 3121 GAGCGTTTGTACCCAGCTCAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
Db 10730 GAGCGTTTGTGTACCCAGCTCAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10789
Qy 3181 CTTCGCGCATGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3240
Db 10790 CTTCGCGCATGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10849
Qy 3241 GACAGAAATGAGCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3300
Db 10850 GACAGAAATGAGCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10909
Qy 3301 GTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3360
Db 10910 GTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10969
Qy 3361 GGCACCGCGGAAAGCGCGCTTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3420
Db 10970 GGCACCGCGGAAAGCGCGCTTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11029
Qy 3421 CGCTGAGGACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3480
Db 11030 CGCTGAGGACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11089
Qy 3481 GTGCTCAATCGCTCTCGGCGAGCTGTATGACAAAGTTTCAATCTCTGCGATGCGAC 3540
Db 11090 GTGCTCAATCGCTCTCGGCGAGCTGTATGACAAAGTTTCAATCTCTGCGATGCGAC 11149
Qy 3541 GGCAGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3600
Db 11150 GGCAGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11209
Qy 3601 CGGTCTGCGCAATCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3660
Db 11210 CGGTCTGCGCAATCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11269
Qy 3661 CGGCGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3720
Db 11270 CGGCGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11329
Qy 3721 ACCCTTCCCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3780
Db 11330 ACCCTTCCCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11389
Qy 3781 GGCAGGAGGAGCATTTGTGGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3840
Db 11390 GGCAGGAGGAGCATTTGTGGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11449
Qy 3841 CGTATTCCAAACGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3900
Db 11450 CGTATTCCAAACGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11509
Qy 3901 CTGCGCTCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3960
Db 11510 CTGCGCTCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11569
```

QY 3961 GTCTCGAGGTGCTCGCAGACCCCGAATCAAGTTCGGCGAGAGCGCTGTTCACCCGC 4020  
DB 11570 GTCTCGAGGTGCTCGCAGACCCCGAATCAAGTTCGGCGAGAGCGCTGTTCACCCGC 11629  
QY 4021 CTGGGCACTGAGTCTGCTCATAGGCGGTGAGCTGGGCAATGATTCGAGGAGGAGCTTAAG 4080  
DB 11630 CTGGGCACTGAGTCTGCTCATAGGCGGTGAGCTGGGCAATGATTCGAGGAGGAGCTTAAG 11689  
QY 4081 CTGAAGCTGTCAGACGATTCCTGTCACAGCTCCCAATATCGCCTGTGTCGCAAAAC 4140  
DB 11690 CTGAAGCTGTCAGACGATTCCTGTCACAGCTCCCAATATCGCCTGTGTCGCAAAAC 11749  
QY 4141 CTGTGAGTCTCTCGCAGACGCTCTCTCTTGGAGCGGTGGCGCGAGAACCTAAGC 4200  
DB 11750 CTGTGAGTCTCTCGCAGACGCTCTCTCTTGGAGCGGTGGCGCGAGAACCTAAGC 11809  
QY 4201 GCAGGCTGCAAAACGATTCCTGTCATCGGCGGCAATCAAGCTGGAAATCATTTGCC 4260  
DB 11810 GCAGGCTGCAAAACGATTCCTGTCATCGGCGGCAATCAAGCTGGAAATCATTTGCC 11869  
QY 4261 CTATGA 4266  
DB 11870 CTATGA 11875

## RESULT 4

US-09-568-480-1

Sequence 1, Application US/09568480

Patent No. 6355458

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zickler, Rose

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

FILE REFERENCE: 4-30582A

CURRENT APPLICATION NUMBER: US/09/568,480

CURRENT FILING DATE: 2000-05-10

PRIOR APPLICATION NUMBER: 09/335,409

PRIOR FILING DATE: 1999-06-17

NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 68750

TYPE: DNA

ORGANISM: Sorangium cellulosum

US-09-568-480-1

Query Match 100.0%; Score 4266; DB 4; Length 68750;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGCGGATGTCCTCCATGAGCGCGACGCCAAGATCCGATTGGGATGTGAGCGAGT 60  
DB 7610 GTGGCGGATGTCCTCCATGAGCGCGACGCCAAGATCCGATTGGGATGTGAGCGAGT 7669  
QY 61 TGGCGTCTGCGCGGGGAGGATGATGTGAGCGGGGTTCTGAGCGCTCTGAGGGGCTCG 120  
DB 7670 TGGCGTCTGCGCGGGGAGGATGATGTGAGCGGGGTTCTGAGCGCTCTGAGGGGCTCG 7729  
QY 121 CGCGACACCGTGGGCGAGTCCCGCGCGAAGCTGGAGTGAAGACGCGTGGTTGATCC 180  
DB 7730 CGCGACACCGTGGGCGAGTCCCGCGCGAAGCTGGAGTGAAGACGCGTGGTTGATCC 7789  
QY 181 GACCCCGATGCGCGGGGAGAGCGCGCTTAAGCGCGCATTTTCTTGAAGCGAGCTAGCC 240  
DB 7790 GACCCCGATGCGCGGGGAGAGCGCGCTTAAGCGCGCATTTTCTTGAAGCGAGCTAGCC 7849  
QY 241 TGGTTCAGCGCTCTTCTTCTGCGATTCGCGCTCGCGAAGCGCTCGGATGAGACCTTCA 300  
DB 241 TGGTTCAGCGCTCTTCTTCTGCGATTCGCGCTCGCGAAGCGCTCGGATGAGACCTTCA 300

DB 7850 TGGTTCAGCGCTCTTCTTCTGCGATTCGCGCTCGCGAAGCGCTCGGATGAGACCTTCA 7909  
QY 301 CATGCACTTCTGTCGAGGTGTGTCGAGGCGCTGAGAGCGCCCGCATTCGTCATCG 360  
DB 7910 CATGCACTTCTGTCGAGGTGTGTCGAGGCGCTGAGAGCGCCCGCATTCGTCATCG 7969  
QY 361 GCGCTGTCGATGCGGAAACGAGAGTTCATCGGATGAGGCGCCGTCGAAATATGAGGCC 420  
DB 7970 GCGCTGTCGATGCGGAAACGAGAGTTCATCGGATGAGGCGCCGTCGAAATATGAGGCC 8029  
QY 421 GCGCTGTCGAGCGAGCGGCGTTCGAGAGATGAGCGCTCATGCGCGGCTGGGAGAGT 480  
DB 8030 GCGCTGTCGAGCGAGCGGCGTTCGAGAGATGAGCGCTCATGCGCGGCTGGGAGAGT 8089  
QY 481 CCCAGGTCGAGCGGCGGAGATTCGTATGCTTGGGCTGGGAGGCGCGTGTTCGCG 540  
DB 8090 CCCAGGTCGAGCGGCGGAGATTCGTATGCTTGGGCTGGGAGGCGCGTGTTCGCG 8149  
QY 541 GTGGATTAAGCGCCATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 600  
DB 8150 GTGGATTAAGCGCCATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8209  
QY 601 TCCGAGGAAATGTCGACAGCGCCCTGAGTGGGATATCGCTATGTTGTCGCGAGACC 660  
DB 8210 TCCGAGGAAATGTCGACAGCGCCCTGAGTGGGATATCGCTATGTTGTCGCGAGACC 8269  
QY 661 CTGCTGTGCTCTGCAAGACCCGCGCGCTGGCGAGGAGCGTGTGCTGCAAGGATTTTCG 720  
DB 8270 CTGCTGTGCTCTGCAAGACCCGCGCGCTGGCGAGGAGCGTGTGCTGCAAGGATTTTCG 8329  
QY 721 GCGGAGGCGATGAGGTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
DB 8330 GCGGAGGCGATGAGGTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8389  
QY 781 AGTGAAGCCCGCGCGAGCGCGATGAGATATGAGGAGTTCGAGGATTCGCGATCAT 840  
DB 8390 AGTGAAGCCCGCGCGAGCGCGATGAGATATGAGGAGTTCGAGGATTCGCGATCAT 8449  
QY 841 CAGGACGCTGCGAGAGCGGCTGATCGCTGCGGAGCGGAGGAGGAGGAGGAGGAGGAGG 900  
DB 8450 CAGGACGCTGCGAGAGCGGCTGATCGCTGCGGAGCGGAGGAGGAGGAGGAGGAGGAGG 8509  
QY 901 AAACGGGCGCTGGGAGAGCGAGGCTGCGCGGCTCTGAGGATATGTCGAGGAGCAC 960  
DB 8510 AAACGGGCGCTGGGAGAGCGAGGCTGCGCGGCTCTGAGGATATGTCGAGGAGCAC 8569  
QY 961 GGCACGGGCAAGACGCTTGTGTAAGCCCATGAATCCAGCTCTGAATGCGGTATACGGC 1020  
DB 8570 GGCACGGGCAAGACGCTTGTGTAAGCCCATGAATCCAGCTCTGAATGCGGTATACGGC 8629  
QY 1021 CTGCGGCGAGATGTCGCAAGCGCGCTGCTGATCGGCTGCTGTAAGACCACTTGGGCGAT 1080  
DB 8630 CTGCGGCGAGATGTCGCAAGCGCGCTGCTGATCGGCTGCTGTAAGACCACTTGGGCGAT 8689  
QY 1081 CTTGATGATGCTGAGGAGTCACTGAGGCTGTAAGGTCGCTTCTTTCAGACAGCGG 1140  
DB 8690 CTTGATGATGCTGAGGAGTCACTGAGGCTGTAAGGTCGCTTCTTTCAGACAGCGG 8749  
QY 1141 CAGATTCTGTCGACCTTCACGCGCAGGCGCTGAACCCCGGATCTCATGGGATGATCTT 1200  
DB 8750 CAGATTCTGTCGACCTTCACGCGCAGGCGCTGAACCCCGGATCTCATGGGATGATCTT 8809  
QY 1201 CGGCTGACCGTCAAGCGCGCGCGCGAGCACGCTGGCGGAGCTGTAAGTACCGCGAGCGG 1260  
DB 8810 CGGCTGACCGTCAAGCGCGCGCGCGAGCACGCTGGCGGAGCTGTAAGTACCGCGAGCGG 8869  
QY 1261 GGGGTGAGCTGTTGCGATGAGGCGGAGACCAACGCGACGCTGCTGTAAGAGAGCGCG 1320  
DB 8870 GGGGTGAGCTGTTGCGATGAGGCGGAGACCAACGCGACGCTGCTGTAAGAGAGCGCG 8929  
QY 1321 GCGGCGAGCTGTCACACCGCGCGCGCGAGCGACGCGAGAGCTGCTGTGTCGCA 1380  
DB 8930 GCGGCGAGCTGTCACACCGCGCGCGCGAGCGACGCGAGAGCTGCTGTGTCGCA 8989

QY 1381 AGGACCGCGTCAGCCCTTGATGACAGAGCGCGCGCTGCGGACCATCTGAGACTTAC 1440  
Db 8990 AGGACCGCGTCAGCCCTTGATGACAGAGCGCGCGCTGCGGACCATCTGAGACTTAC 9049  
QY 1441 CTTTGGCAGTGTCTGGGCGCATGTGGCGTTCACTGTGGCGACACGCGCGCATGTGAG 1500  
Db 9050 CTTTGGCAGTGTCTGGGCGCATGTGGCGTTCACTGTGGCGACACGCGCGCATGTGAG 9109  
QY 1501 CACCGGCTCGCGGTGCGCGCGCGACGTCGAGAGAGGGGCTGCGGCGACGCTGGACGCTGCG 1560  
Db 9110 CACCGGCTCGCGGTGCGCGCGCGACGTCGAGAGAGGGGCTGCGGCGACGCTGGACGCTGCG 9169  
QY 1561 GCGCAGGAGCAGACGTCGCGCGCGGTGCGCGCGCATGTGCGCATTTCTCACGCGGCAAG 1620  
Db 9170 GCGCAGGAGCAGACGTCGCGCGCGGTGCGCGCGCATGTGCGCATTTCTCACGCGGCAAG 9229  
QY 1621 CTCGCTTTCTTTTCAACGCGACAGAGGCGCGACAGCGTGGGCAATGGGCGTGGCTTAC 1680  
Db 9230 CTCGCTTTCTTTTCAACGCGACAGAGGCGCGACAGCGTGGGCAATGGGCGTGGCTTAC 9289  
QY 1681 GATGTATGGTCCGCGTTCCGCGAGGCGTTGCACTGTGCTGAGGCTGTTCAACGAGAG 1740  
Db 9290 GATGTATGGTCCGCGTTCCGCGAGGCGTTGCACTGTGCTGAGGCTGTTCAACGAGAG 9349  
QY 1741 CTCGACCGGCGCTCCGCGAGGTGATGTGGCGCGAACCGGCGAGGCTGACGCGCGCTG 1800  
Db 9350 CTCGACCGGCGCTCCGCGAGGTGATGTGGCGCGAACCGGCGAGGCTGACGCGCGCTG 9409  
QY 1801 CTCGACCGAGCAGCTTTCAACCGACGCGCGCTGTTTCACTTGAATATGCTGCGCGCG 1860  
Db 9410 CTCGACCGAGCAGCTTTCAACCGACGCGCGCTGTTTCACTTGAATATGCTGCGCGCG 9469  
QY 1861 CTGTGCGGCTGTGGGGGTGTAGAGCGGAGTTGTGCGCGCGCATATGATGCTGTAGCTG 1920  
Db 9470 CTGTGCGGCTGTGGGGGTGTAGAGCGGAGTTGTGCGCGCATATGATGCTGTAGCTG 9529  
QY 1921 GTGGCTGCTGCTGCGTGGCGCGCGCTGTTTCTGCTTGAAGACGCGGTTCTGTGGCTGCG 1980  
Db 9530 GTGGCTGCTGCTGCGTGGCGCGCGCTGTTTCTGCTTGAAGACGCGGTTCTGTGGCTGCG 9589  
QY 1981 CGCGGCGCGCTGTATGACAGCGCTGCGCGCGCGCGCGCGCATGTGTGATGACGCGCGCG 2040  
Db 9590 CGCGGCGCGCTGTATGACAGCGCTGCGCGCGCGCGCGCGCATGTGTGATGACGCGCGCG 9649  
QY 2041 GAGGCGGATGTGGCTGTGCGGTGGCGCGCGACGAGGCTGTGCTGATGCTGCGCGCGCTG 2100  
Db 9650 GAGGCGGATGTGGCTGTGCGGTGGCGCGCGACGAGGCTGTGCTGATGCTGCGCGCGCTG 9709  
QY 2101 AACGCTCGGACACAGGTGTGATGCGCGGCGCGCGGCAACCGCTGATGCTGATGCGGCG 2160  
Db 9710 AACGCTCGGACACAGGTGTGATGCGCGGCGCGCGGCAACCGCTGATGCTGATGCGGCG 9769  
QY 2161 GCGATGCGCGCGCGCGCGCGCGGAAACAAAGCGCTTCCAGTCTGCGATGCTTCACTCA 2220  
Db 9770 GCGATGCGCGCGCGCGCGCGCGGAAACAAAGCGCTTCCAGTCTGCGATGCTTCACTCA 9829  
QY 2221 CCGCTCATGCGCGCGCGCGCGCGGAGCGTTCCGCGCGTGTGGCGCGATGCTGATGCTACCGG 2280  
Db 9830 CCGCTCATGCGCGCGCGCGCGCGGAGCGTTCCGCGCGTGTGGCGCGATGCTGATGCTACCGG 9889  
QY 2281 CCGCGCTGATGCTGCTGCTGATGCAATCTGACGCGGAGGCTTTCACAGACAGAGGTGAGC 2340  
Db 9890 CCGCGCTGATGCTGCTGCTGATGCAATCTGACGCGGAGGCTTTCACAGACAGAGGTGAGC 9949  
QY 2341 TCGCGCGCTATTTGGGTGCGCAACGCGCGAGAGGTGTGCGCTTTCGCGGATGAGTGAAG 2400  
Db 9950 TCGCGCGCTATTTGGGTGCGCAACGCGCGAGAGGTGTGCGCTTTCGCGGATGAGTGAAG 10009  
QY 2401 GCGCTGACAGCGCGCGGTGCGCGGACCTTGTGAGGTGTGCTGCGGATGAGTGAAG 2460  
Db 10010 GCGCTGACAGCGCGCGGTGCGCGGACCTTGTGAGGTGTGCTGCGGATGAGTGAAG 10069

QY 2461 GCGCTGTGCTGCTGCTGATGCGGACCGCGCGCGCGCGCTGCTGATGCTGCGCGCT 2520  
Db 10070 GCGCTGTGCTGCTGCTGATGCGGACCGCGCGCGCGCGCTGCTGATGCTGCGCGCT 10129  
QY 2521 GGGCGGTGACGAGCGCGGACCGGTGTGAGAGCGCTGCGCGCGCTTGGGCGCTGCGTGGC 2580  
Db 10130 GGGCGGTGACGAGCGCGGACCGGTGTGAGAGCGCTGCGCGCGCTTGGGCGCTGCGTGGC 10189  
QY 2581 CTGTCTCTGCGGCGCGCTCTTCCCTCAGGGGGGCGCGGCTGCGCTGCGGACGTTAC 2640  
Db 10190 CTGTCTCTGCGGCGCGCTCTTCCCTCAGGGGGGCGCGGCTGCGCTGCGGACGTTAC 10249  
QY 2641 CTTTGGCAGCGGAGCGCTTATGATGACACGAAAGCTTCAACGCGCGCGCTGCGGAC 2700  
Db 10250 CTTTGGCAGCGGAGCGCTTATGATGACACGAAAGCTTCAACGCGCGCGCTGCGGAC 10309  
QY 2701 CGCGGTGCTTCCGGAGCGGGTCAACGAGGTTCAGAGAGGGGGGCGGTGCGCGCGCG 2760  
Db 10310 CGCGGTGCTTCCGGAGCGGGTCAACGAGGTTCAGAGAGGGGGGCGGTGCGCGCGCG 10369  
QY 2761 GACCGGCGGAGCGCTCGGCTGACCATCGCGCGCGGAGAGCGCGCGGAGAAAGTTC 2820  
Db 10370 GACCGGCGGAGCGCTCGGCTGACCATCGCGCGCGGAGAGCGCGCGGAGAAAGTTC 10429  
QY 2821 GAGGCGCGGAGCGCGCTCGGCTCGGAGTGAATGAGCCAGGCGCTGTTGATCAC 2880  
Db 10430 GAGGCGCGGAGCGCGCTCGGCTCGGAGTGAATGAGCCAGGCGCTGTTGATCAC 10489  
QY 2881 CTCGTGCTTCCGGGTGACGAGAGCGCGCGCGCTTGTGGGCGAGTGAATGCGCGCTC 2940  
Db 10490 CTCGTGCTTCCGGGTGACGAGAGCGCGCGCGCTTGTGGGCGAGTGAATGCGCGCTC 10549  
QY 2941 GACGCGGCGGCGCTCAGCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 3000  
Db 10550 GACGCGGCGGCGCTCAGCTTCAATGATGATGATGATGATGATGATGATGATGATG 10609  
QY 3001 CTGCGGAGAAAGCCCAACCTTCCGCTGCTGCTGAGGCGAGTGGCGCGCATGCTC 3060  
Db 10610 CTGCGGAGAAAGCCCAACCTTCCGCTGCTGCTGAGGCGAGTGGCGCGCATGCTC 10669  
QY 3061 GCGGTGGCGAGGGGTGATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3120  
Db 10670 GCGGTGGCGAGGGGTGATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10729  
QY 3121 GAGGCTTGTGCTAACCAAGTCAACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3180  
Db 10730 GAGGCTTGTGCTAACCAAGTCAACGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10789  
QY 3181 CTCTCGGAGATGAGGCGCGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3240  
Db 10790 CTCTCGGAGATGAGGCGCGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10849  
QY 3241 GACGAGATAGCCCGCTTCAAGCGGAGGAGCGGCTGCTGATTCATGCGGCGACCGCGCG 3300  
Db 10850 GACGAGATAGCCCGCTTCAAGCGGAGGAGCGGCTGCTGATTCATGCGGCGACCGCGCG 10909  
QY 3301 GTGCGTCTGCGCGGCTGACAGTGGCGGAGACGAGGAGCGGAGTCCATGCGGCGCG 3360  
Db 10910 GTGCGTCTGCGCGGCTGACAGTGGCGGAGACGAGGAGCGGAGTCCATGCGGCGCG 10969  
QY 3361 GGCAGCGCGCGAGAAACGCGCTTACCTGAGTCTGCTGCGGCTGCTGCTGCTGCTGCTG 3420  
Db 10970 GGCAGCGCGCGAGAAACGCGCTTACCTGAGTCTGCTGCGGCTGCTGCTGCTGCTGCTG 11029  
QY 3421 CGCTTGAACCGGCTTCTGCGGAGCGTGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 3480  
Db 11030 CGCTTGAACCGGCTTCTGCGGAGCGTGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 11089  
QY 3481 GTGCTCAATGCTGCTGCGGAGGAGTGAATGACAAAGTTCATCTCTGCGATGCGAC 3540  
Db 11090 GTGCTCAATGCTGCTGCGGAGGAGTGAATGACAAAGTTCATCTCTGCGATGCGAC 11149  
QY 3541 GCGCGGTTTGTGAGCTTGGAGGCGAGCTGTTACGCGGATTAACAGCTGCGGCTGCGG 3600

Db	11150	GGCCGCGTTTGTGAGAGCTCGGCAAGCGGAGCTGTTTACCGGATTAACGAGCTCGGGCTCGGG	11209
OY	3601	CCGTTTCTGGCGCAATCTCTCTCTCTCTGCGTGGTGATCTCCGGGGATGATGCTCGACGG	3660
Db	11210	CCGTTTCTGGCGCAATCTCTCTCTCTCTGCGTGGTGATCTCCGGGGATGATGCTCGACGG	11268
OY	3661	CCGGCGGGGGGTCGTTGGGCTCTTGGAGAGAGCTCTCGGGCCCTGATTCGAGAGGCGGTTC	3720
Db	11270	CCGGCGGGGGGTCGTTGGGCTCTTGGAGAGAGCTCTCGGGCCCTGATTCGAGAGGCGGTTC	11328
OY	3721	ACCCCTCCCCCATCGGAGCGCTCCGATCGCCGTGTGCTGCGCATGCTTCGAGCATG	3780
Db	11330	ACCCCTCCCCCATCGGAGCGCTCCGATCGCCGTGTGCTGCGCATGCTTCGAGCATG	11389
OY	3781	GCGAGGCGCGAGCATCTTGGGAGAGCTGTACTACAGCTGGGTGATCCCGAGAGTTCAGATC	3840
Db	11390	GCGAGGCGCGAGCATCTTGGGAGAGCTGTACTACAGCTGGGTGATCCCGAGAGTTCAGATC	11449
OY	3841	CGTATTCACACCCACGAGAGGCGCGGCGCGTCCACCGGGGATCGGGAGCTGCTCGACAG	3900
Db	11450	CGTATTCACACCCACGAGAGGCGCGGCGCGTCCACCGGGGATCGGGAGCTGCTCGACAG	11509
OY	3901	CTCGCGTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCTCTCGTACGAG	3960
Db	11510	CTCGCGTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCTCTCGTACGAG	11569
OY	3961	GTCTTCGAGAGTGTCTGGCAGCGCCCGGAATAGATGTGGCGCGGAGCGCTGTTCACCCGC	4020
Db	11570	GTCTTCGAGAGTGTCTGGCAGCGCCCGGAATAGATGTGGCGCGGAGCGCTGTTCACCCGC	11629
OY	4021	CTCGGCAATGAGCTCGCTCATAGCGCGTGGAGCTGGCGAATGTATCGAGGCGAGCTCTAAG	4080
Db	11630	CTCGGCAATGAGCTCGCTCATAGCGCGTGGAGCTGGCGAATGTATCGAGGCGAGCTCTAAG	11689
OY	4081	CTGAAGCTGTGACGAGCGTTCTGTGCAAGTCCCCCAATATGCGCTTGTGGCCCAAAAC	4140
Db	11690	CTGAAGCTGTGACGAGCGTTCTGTGCAAGTCCCCCAATATGCGCTTGTGTGGCCCAAAAC	11749
OY	4141	CTGTTGAGATGCTCTCGGCAAGAGCTCTCTCTTGGAGGGGGTGGCGGCGAGAACTTAAG	4200
Db	11750	CTGTTGAGATGCTCTCGGCAAGAGCTCTCTCTTGGAGGGGGTGGCGGCGAGAACTTAAG	11809
OY	4201	GCAGGCGGTGCAAAACGACTTGTCTCATCGGCGCGAGATCAAGACTGGAAATCATTTGCC	4260
Db	11810	GCAGGCGGTGCAAAACGACTTGTCTCATCGGCGCGAGATCAAGACTGGAAATCATTTGCC	11869
OY	4261	CTATGA 4266	
Db	11870	CTATGA 11875	
RESULT 5			
US-09-568-486-1			
; Sequence 1, Application US/09568486			
; Patent No. 6355459			
; GENERAL INFORMATION:			
; APPLICANT: Schupp, Thomas			
; APPLICANT: Ligon, James			
; APPLICANT: Molnar, Istvan			
; APPLICANT: Zinkle, Ross			
; APPLICANT: Cyr, Devon			
; APPLICANT: Goerlach, Joern			
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES			
; FILE REFERENCE: 4-30582A			
; CURRENT APPLICATION NUMBER: US/09/568,486			
; CURRENT FILING DATE: 2000-05-10			
; PRIOR APPLICATION NUMBER: 09/335,409			
; PRIOR FILING DATE: 1999-06-17			
; NUMBER OF SEQ ID NOS: 30			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 1			
; LENGTH: 68750			

	Query Match	100.0%	Score 4266;	DB 4;	Length 68750;
	Best Local Similarity	100.0%	Pident. No. 0;	Mismatches 0;	Indels 0;
	Matches 4266;	Conservative 0;	Prd. Matches 0;	Indels 0;	Gaps 0;
US-09-568-486-1					
TYPE: DNA					
ORGANISM: Sorangium cellulosum					
1	GTGCGGATCGTCCCATCGACGCGCGCAAGATCCGATTGCGATGTCGAGCGAGT	60			
Db	7610 GTGCGGATCGTCCCATCGACGCGCGCAAGATCCGATTGCGATGTCGAGCGAGT	7669			
Qy	61 TGCCGTCGCCCCGATGATGATCTGAGCGGATTGAGACCTTCCTGAGGCTCG	120			
Db	7670 TGCCGTCGCCCCGATGATGATCTGAGCGGATTGAGACCTTCCTGAGGCTCG	7729			
Qy	121 CGGCAACCGTCGGGCGAGTCCCGCGGAAACGCTGGGATGAGAGGATGATGATCC	180			
Db	7730 CGGCAACCGTCGGGCGAGTCCCGCGGAAACGCTGGGATGAGAGGATGATGATCC	7789			
Qy	181 GACCCCGATGCCCCGGGGAAGACGCCCCGTTACGCGCATCTTCTGAGCGATGACC	240			
Db	7790 GACCCCGATGCCCCGGGGAAGACGCCCCGTTACGCGCATCTTCTGAGCGATGACC	7849			
Qy	241 TGCTTCGACGCTCTCTTCTTGCGATTTGCTTCGCAAGCGCTGCGGATGACCCGCA	300			
Db	7850 TGCTTCGACGCTCTCTTCTTGCGATTTGCTTCGCAAGCGCTGCGGATGACCCGCA	7909			
Qy	301 CATGACCTCTTGCTGAGAGTGTGCTGGGAGCGCTGAGAGACGCGCGATCGCTCATCG	360			
Db	7910 CATGACCTCTTGCTGAGAGTGTGCTGGGAGCGCGCTGAGAGACGCGCGATCGCTCATCG	7969			
Qy	361 GCCTTCCTCGATCGAAGACGCGAGTGTTCATCGGATCGGCGCTCCGAATATGAGCC	420			
Db	7970 GCCTTCCTCGATCGAAGACGCGAGTGTTCATCGGATCGGCGCTCCGAATATGAGCC	8029			
Qy	421 GCGTCGCGCAAGGAGCGGCGTCCGAGAGATCGACGCTCATGCGGCGTGGGACGATG	480			
Db	8030 GCGTCGCGCAAGGAGCGGCGTCCGAGAGATCGACGCTCATGCGGCGTGGGACGATG	8089			
Qy	481 CCGACGCTCGAGGCGGCGCGAATCTCGATGCCCCCTCGGCGCTGCGAGGCGCTGTGCGG	540			
Db	8090 CCGACGCTCGAGGCGGCGCGAATCTCGATGCCCCCTCGGCGCTGCGAGGCGCTGTGCGG	8149			
Qy	541 GTGATATGCGCTATTTGCTCTCGCTGTGCGCTTATCTGCGCTTCAGAGTTTCCG	600			
Db	8150 GTGATATGCGCTATTTGCTCTCGCTGTGCGCTTATCTGCGCTTCAGAGTTTCCG	8209			
Qy	601 TCCGGGATGCTCCACGCGCTTGCTGCTGAGGATGCTGATGTTGTGCGCCAGCAC	660			
Db	8210 TCCGGGATGCTCCACGCGCTTGCTGCTGAGGATGCTGATGTTGTGCGCCAGCAC	8269			
Qy	661 CTGCTGTGCTCTTGAAGATCCGGGCGCTGCGCAGGACGCTGCTGCAAGGATTTTCC	720			
Db	8270 CTGCTGTGCTCTTGAAGATCCGGGCGCTGCGCAGGACGCTGCTGCAAGGATTTTCC	8329			
Qy	721 GCGGAGGCGGATGAGTTCGAGCAGAGGAGGAGCGCCGCTGCTGCTTCAGACGCTC	780			
Db	8330 GCGGAGGCGGATGAGTTCGAGCAGAGGAGGAGCGCCGCTGCTGCTTCAGACGCTC	8389			
Qy	781 AGTGAAGCCGCGCGAGCGCGATCGATATTTGGCGGTGATTGAGGATCCGCATCAAT	840			
Db	8390 AGTGAAGCCGCGCGAGCGCGATCGATATTTGGCGGTGATTGAGGATCCGCATCAAT	8449			
Qy	841 CACGACGCTGGAACAGCGGTCTGACCGGTGCGAAGCGGAGCTCCCAAGAAATGCTGTG	900			
Db	8450 CACGACGCTGGAACAGCGGTCTGACCGGTGCGAAGCGGAGCTCCCAAGAAATGCTGTG	8509			
Qy	901 AAACGGGCGCTGAGGAGCGCAGGCTGCGCGCTCTTCTGCTGAGTTATGTCGAGCAC	960			
Db	8510 AAACGGGCGCTGAGGAGCGCAGGCTGCGCGCTCTTCTGCTGAGTTATGTCGAGCAC	8569			
Qy	961 GGACGCGGACGACGCTTGATGACCCCATCGAAATCCAACTCTGATTCGGTATACGCG	1020			

Db	8570	GGCAGCGGCCACGACCCCTTGCTGACCCCATCGAATTCAGACTCTGAATGCGGTATACGCG	86229
Qy	1021	CTCGGGCGAGATGTCGCGCACGCGCGCTGCTGATCGGGTCGGTGAAGAACCAACCTTGGCCAT	1080
Db	8630	CTCGGGCGAGATGTCGCGCACGCGCGCTGCTGATCGGGTCGGTGAAGAACCAACCTTGGCCAT	8689
Qy	1081	CCTGAGTATGCGTCGGGGATCACTGCGCTGCTGAAGGTCGTCCTTCCCTTACAGACGGG	1140
Db	8690	CCTGAGTATGCGTCGGGGATCACTGCGCTGCTGAAGGTCGTCCTTCCCTTACAGACGGG	8749
Qy	1141	CAGATTCCTMGCGCACCTCACGCGCGAGGCGGTGAACCCCGGATCTCATGGGGTGATCTT	1200
Db	8750	CAGATTCCTMGCGCACCTCACGCGCGAGGCGGTGAACCCCGGATCTCATGGGGTGATCTT	8809
Qy	1201	CGGCTGACCGGTACACGCGCGCGCCCGGACACCGGTGCGCGGACTGGAATACCCCGCGACGGCG	1260
Db	8810	CGGCTGACCGGTACACGCGCGCGCCCGGACACCGGTGCGCGGACTGGAATACCCCGCGACGGCG	8869
Qy	1261	GCGGTGACCTCGTTCGGCATGAGCGCGGACCAACGCGCACGTTGTCGTTGAAGAGCGCGCG	1320
Db	8870	GCGGTGACCTCGTTCGGCATGAGCGCGGACCAACGCGCACGTTGTCGTTGAAGAGCGCGCG	8929
Qy	1321	GCGGGAAGTGTCCACACGCGCGCGCGCGCGGACCGCGAGACCTGCTGCTGTCCTGTCGCA	1380
Db	8930	GCGGGAAGTGTCCACACGCGCGCGCGCGCGGACCGCGAGACCTGCTGCTGTCCTGTCGCA	8989
Qy	1381	AGGACCGCGGTACAGCCCTGGAATGACAGCGCGCGCGCGCTGCGCGACATCTGGAAGCTTAC	1440
Db	8990	AGGACCGCGGTACAGCCCTGGAATGACAGCGCGCGCGCGCTGCGCGACATCTGGAAGCTTAC	9049
Qy	1441	CTTTCGACGTGTCTGGGCGGATGTGGCGTTTCACTGTCGAGACGCGCGAGCGCGATGAG	1500
Db	9050	CTTTCGACGTGTCTGGGCGGATGTGGCGTTTCACTGTCGAGACGCGCGAGCGCGATGAG	9109
Qy	1501	CACCGGCTTCGGGGTGGCGCGGACGCTCGAGGGAGGGGGCTGCGGGACAGCCTTGAAGCTTGC	1560
Db	9110	CACCGGCTTCGGGGTGGCGCGGACGCTCGAGGGAGGGGGCTGCGGGACAGCCTTGAAGCTTGC	9169
Qy	1561	GCGCAGGGACAGACGTCGCGCGCGGTGCGGGTGGCGAGTATCGCGGATCTCTCACGCGGCAAG	1620
Db	9170	GCGCAGGGACAGACGTCGCGCGCGGTGCGGGTGGCGAGTATCGCGGATCTCTCACGCGGCAAG	9229
Qy	1621	CTCGGCTTCTCTTCAACCGGACAGGGGGCGGAGCGCTGGGCATGTGGCCGTGGCTGTAC	1680
Db	9230	CTCGGCTTCTCTTCAACCGGACAGGGGGCGGAGCGCTGGGCATGTGGCCGTGGCTGTAC	9289
Qy	1681	GATGTAATGTCGCGGTTCCGCGAGGCGTTTCCACTGTGCGTAGGCTTTCACACGAGAG	1740
Db	9290	GATGTAATGTCGCGGTTCCGCGAGGCGTTTCCACTGTGCGTAGGCTTTCACACGAGAG	9349
Qy	1741	CTCGACCGGGCGGCTCCGCGAGGTGATGTGGGCCGAACCGGCGACGTCGAGACGCGCGCTG	1800
Db	9350	CTCGACCGGGCGGCTCCGCGAGGTGATGTGGGCCGAACCGGCGACGTCGAGACGCGCGCTG	9409
Qy	1801	CTCGACCGAGACGCTTCAACCGACGCGCGCGCTGTTCACCTTCGAATATGCGCTCGCGCG	1860
Db	9410	CTCGACCGAGACGCTTCAACCGACGCGCGCGCTGTTCACCTTCGAATATGCGCTCGCGCG	9469
Qy	1861	CTGTGGCGGTGCTGTGGGGTGTAGAGCGCGGAGTTTGTCGCGCGCCATAGCATGGTGAAGCTG	1920
Db	9470	CTGTGGCGGTGCTGTGGGGTGTAGAGCGCGGAGTTTGTCGCGCGCCATAGCATGGTGAAGCTG	9529
Qy	1921	GTCGCTGCTCCTGTCGCGCGCGCGCTGTTCTTCGCTTGAAGACGCGGCTTCTTGGTGGCTGCG	1980
Db	9530	GTCGCTGCTCCTGTCGCGCGCGCGCTGTTCTTCGCTTGAAGACGCGGCTTCTTGGTGGCTGCG	9589
Qy	1981	CGCGGGCCCTGATGACAGGCGCTGCGCGGCGCGGGCGGATGATGTCATGAGAGCGCGCG	2040
Db	9590	CGCGGGCCCTGATGACAGGCGCTGCGCGGCGCGGGCGGATGATGTCATGAGAGCGCGCG	9649
Qy	2041	GAGGCGGATGTGCTGTCGCGGTGGCGCGGACGCGAGCTCGGTTCGATGCGCGCGCTC	2100

Db	9650	GAGGCGCATATGAGTGTCTGTGCGGTGGGCGCGACAGACGATCGAGTTCGATTCGCGCGGCTC	9709
QY	2101	AAGCCTCCGGAACCAAGGTGTATCGGCGGCGCGCGGCAACCGGTGATCGATCGCGCGC	2160
Db	9710	AACGCTCCGGAACCAAGGTGTATCGGCGGCGCGCGGCAACCGGTGATCGATCGCGCGC	9769
QY	2161	GCATATGACCCCGCGCGGGGCGGAAACCAAGCGCTCAAGTCTCGGATCGTTCCACTCA	2220
Db	9770	GCATATGACCCCGCGCGGGGCGGAAACCAAGCGCTCAAGTCTCGGATCGTTCCACTCA	9829
QY	2221	CCGCTCATATGAGCCCGGAATGCTGAGAGCGTTGCGGCGGTGGGCGAGTCCGATGAGCTACCGG	2280
Db	9830	CCGCTCATATGAGCCCGGAATGCTGAGAGCGTTGCGGCGGTGGGCGAGTCCGATGAGCTACCGG	9889
QY	2341	TCGCGCGGACTATTTGGGTGTGCGCCACGCGCGAGAGGTGTGCGCTTCGCGGATGGAGTGAAG	2400
Db	9950	TCGCGCGGACTATTTGGGTGTGCGCCACGCGCGAGAGGTGTGCGCTTCGCGGATGGAGTGAAG	10090
QY	2401	GGCGTGAACGCGCGCGGTGGGGGCACTTCGTGAGAGTGGATCGGAATCGACGCGTGC	2460
Db	10010	GGCGTGAACGCGCGCGGTGGGGGCACTTCGTGAGAGTGGATCGGAATCGACGCGTGC	10068
QY	2461	GGCCTGGTGGCTGTGCTGATCGCGGACGCGCGCGCGCTGTGCTGATCGTGC	2520
Db	10070	GGCCTGGTGGCTGTGCTGATCGCGGACGCGCGCGCGCGCTGTGCTGATCGTGC	10128
QY	2521	GGGCGTGAACGACCGCGGACCGTGTGAGAGCGCTTGGCGGCTCTTGGGCGGTGGGCG	2580
Db	10130	GGGCGTGAACGACCGCGGACCGTGTGAGAGCGCTTGGCGGCTCTTGGGCGGTGGGCG	10188
QY	2581	CTGGTCTCTGGGGCGGCGCTCTTCCCTCAAGGGGGGGCGGGTGGCGGTGCCCAAGTAC	2640
Db	10190	CTGGTCTCTGGGGCGGCGCTCTTCCCTCAAGGGGGGGCGGGTGGCGGTGCCCAAGTAC	10248
QY	2641	CCTTGGACGAGCGGACGCTTCTGATTCGACACCAAAAGCCGACGACGCGCGCTGGCGAC	2700
Db	10250	CCTTGGACGAGCGGACGCTTCTGATTCGACACCAAAAGCCGACGACGCGCGCTGGCGAC	10308
QY	2701	CGCGTGTCTCGGAGCGGGTCAACGACGAGTTCGAGAGGGGGGCGCGGTGCGCGCGGCG	2760
Db	10310	CGCGTGTCTCGGAGCGGGTCAACGACGAGTTCGAGAGGGGGGCGCGGTGCGCGCGGCG	10368
QY	2761	GACCGGCGGACGCTGTGGGCTTCGACATCCGCGCGCCCGGAGAGCGGAAACCGGAGAAAGTTC	2820
Db	10370	GACCGGCGGACGCTGTGGGCTTCGACATCCGCGCGCCCGGAGAGCGGAGAAAGTTC	10428
QY	2821	GAGGCGCGCGGACGCTCGGTTCCGAGTGAATGATGAGCGAGCGGTGATTCAC	2880
Db	10430	GAGGCGCGCGGACGCTCGGTTCCGAGTGAATGATGAGCGAGCGGTGATTCAC	10488
QY	2881	CTCGTGTTCGAGGTCAACGAGCGGCGCGCCCTGTGTCTGGGCGAGGTTCGACATCGCGCTC	2940
Db	10490	CTCGTGTTCGAGGTCAACGAGCGGCGCGCCCTGTGTCTGGGCGAGGTTCGACATCGCGCTC	10548
QY	2941	GACCGGCGGAGGCTCAGCTTCATATGATTCAGCTGCGCTGGGCAATGTTGCCCGACGAC	3000
Db	10550	GACCGGCGGAGGCTCAGCTTCATATGATTCAGCTGCGCTGGGCAATGTTGCCCGACGAC	10608
QY	3001	CTGCGCGGAAAGCCCAACCTTCGCGGTGTCTGAGAGGGAGTGGCGCGGGGCGACATCGTC	3060
Db	10610	CTGCGCGGAAAGCCCAACCTTCGCGGTGTCTGAGAGGGAGTGGCGCGGGGCGACATCGTC	10668
QY	3061	GCCGTGGGCGAGGCGGTGAACGCGCTCTGTGTGGGCGCAACCGGTCAATCGCCCTTTCGAGCG	3120
Db	10670	GCCGTGGGCGAGGCGGTGAACGCGCTCTGTGTGGGCGCAACCGGTCAATCGCCCTTTCGAGCG	10728
QY	3121	GGAGCGTTTGTCTACCGACGTCAACGATCGATCGGCTGTGCTGTGCTCTGAGCTCAGAGCG	3180
Db	10730	GGAGCGTTTGTCTACCGACGTCAACGATCGATCGGCTGTGCTGTGCTCTGAGCTCAGAGCG	10788





601 TCCGGGATATCTCCACGGCCCTGGCTGGGGATATCGCTGATTTGTTCGCCAGACACC 660  
Db TCCGGGATATCTCCACGGCCCTGGCTGGGGATATCGCTGATTTGTTCGCCAGACACC 8269  
Qy CTCTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGAGAACGGTCCGTGCAAGGCAATTTTCG 720  
Db CTCTGTGGCTCTCGAAGACCCGGGCGCTGGCCAGAGAACGGTCCGTGCAAGGCAATTTTCG 8329  
Qy GCGGAGCCGATATGGGTTGGAGCGAGCGAAGGGTGGCGCTGCTGATTCCTCAAGGGGCTC 780  
Db GCGGAGCCGATATGGGTTGGAGCGAGCGAAGGGTGGCGCTGCTGATTCCTCAAGGGGCTC 8389  
Qy AGTGAAGCCCGCGCGAGCGAGCGATATTTGGCGGTGATTTCGAGGATTCGCGATCAAT 840  
Db AGTGAAGCCCGCGCGAGCGAGCGATATTTGGCGGTGATTTCGAGGATTCGCGATCAAT 8449  
Qy CACGACGGTGGAGCAGCGGTCTGACCGTCCGAAACGGAGCTCCCAAGAAATCGTCTG 900  
Db CACGACGGTGGAGCAGCGGTCTGACCGTCCGAAACGGAGCTCCCAAGAAATCGTCTG 8509  
Qy AAACGGGCGCTGGCGGAGCGAGCGCTGGCTGCTGATGGGTTATGCGAGGACAC 960  
Db AAACGGGCGCTGGCGGAGCGAGCGCTGGCTGCTGATGGGTTATGCGAGGACAC 8569  
Qy GGCACGGGACGACGCTTGTGACCCCATCGAAATCCAACTCTGAATGCGGTATACGGC 1020  
Db GGCACGGGACGACGCTTGTGACCCCATCGAAATCCAACTCTGAATGCGGTATACGGC 8629  
Qy CTCTGGGCGAATATGTCGACCGCGCTGCTGATCGGGTGGTGAAGACCAACTTGGCAT 1080  
Db CTCTGGGCGAATATGTCGACCGCGCTGCTGATCGGGTGGTGAAGACCAACTTGGCAT 8689  
Qy CTCTGATATGGTGGGGGATCATCTGGGCTGCTGAAGTCTCTTTCCTTCAGACGG 1140  
Db CTCTGATATGGTGGGGGATCATCTGGGCTGCTGAAGTCTCTTTCCTTCAGACGG 8749  
Qy CAGATTCTCTGGGCACTTCACCGCGAGGCGCTGAACCCCGGATCTCAATGGGGTATCTT 1200  
Db CAGATTCTCTGGGCACTTCACCGCGAGGCGCTGAACCCCGGATCTCAATGGGGTATCTT 8809  
Qy CGGCTGACCTGTCACCGCGCGCGGACACCTGTCGCGAATGAAATCCCGCGAGCGGCG 1260  
Db CGGCTGACCTGTCACCGCGCGCGGACACCTGTCGCGAATGAAATCCCGCGAGCGGCG 8869  
Qy GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAAGGCGCG 1320  
Db GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAAGGCGCG 8929  
Qy GCGGCGAGTGCACACCGCGCGCGCGAGCGACCGGACAGCTCTGCTGTCGCGCA 1380  
Db GCGGCGAGTGCACACCGCGCGCGCGAGCGACCGGACAGCTCTGCTGTCGCGCA 8989  
Qy AGGACCGGCTACCCCTGATGACAGGCGGGCGGCTGCGGACCATCTGAGACCTAC 1440  
Db AGGACCGGCTACCCCTGATGACAGGCGGGCGGCTGCGGACCATCTGAGACCTAC 9049  
Qy CTTTTCGAGTGTCTGGGCGATGTCGCTTCACTCTGGGCGACGACCGCGATGAG 1500  
Db CTTTTCGAGTGTCTGGGCGATGTCGCTTCACTCTGGGCGACGACCGCGATGAG 9109  
Qy CACCGGCTCGCGGTCGCGGCGACGTCGAGAGGAGGGGCTGGGCGACCTGAGCGCTCG 1560  
Db CACCGGCTCGCGGTCGCGGCGACGTCGAGAGGAGGGGCTGGGCGACCTGAGCGCTCG 9169  
Qy GCGCAGGAGCAGACGTCGCGCGGTCGCGGTCGCGATTCGATTCGACGCGGCAAG 1620  
Db GCGCAGGAGCAGACGTCGCGCGGTCGCGGTCGCGATTCGATTCGACGCGGCAAG 9229  
Qy CTCGCTTCTCTTCAACGAGGAGGCGAGACGTCGGGATGAGGCTGAGCTGAC 1680  
Db CTCGCTTCTCTTCAACGAGGAGGCGAGACGTCGGGATGAGGCTGAGCTGAC 9289

Qy GATGATATGTCGCGCTTCGCGAGCGTTGCACTGTGCGTGAAGCTGTTCAACGAGAG 1740  
Db GATGATATGTCGCGCTTCGCGAGCGTTGCACTGTGCGTGAAGCTGTTCAACGAGAG 9349  
Qy CTGCAACCGCGCTTCGCGAGGATGATGTGGCCGAACTGGGCGAGCTGAGACCGCGCTG 1800  
Db CTGCAACCGCGCTTCGCGAGGATGATGTGGCCGAACTGGGCGAGCTGAGACCGCGCTG 9409  
Qy CTGCAACGACAGGCTTCACCGCGAGCGGCTGTTCACTTCGAAATATGCGCTGCGCG 1860  
Db CTGCAACGACAGGCTTCACCGCGAGCGGCTGTTCACTTCGAAATATGCGCTGCGCG 9469  
Qy CTGTGGCGGTCGTGGGTTGTAAGCCGAGATTGATGCGCGGCAATAGCATCGTGAAGCTG 1920  
Db CTGTGGCGGTCGTGGGTTGTAAGCCGAGATTGATGCGCGGCAATAGCATCGTGAAGCTG 9529  
Qy GTGGCTGCTGCGTGGCGGCGGTGTTCTGCTTGAAGACGCGGTGTTCTGTGGCTGCG 1980  
Db GTGGCTGCTGCGTGGCGGCGGTGTTCTGCTTGAAGACGCGGTGTTCTGTGGCTGCG 9589  
Qy CGCGGCGCGCTGATGACAGCGCTGCGCGGCGCGGCGGAGATGATGATGAGGCGCG 2040  
Db CGCGGCGCGCTGATGACAGCGCTGCGCGGCGCGGCGGAGATGATGATGAGGCGCG 9649  
Qy GAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100  
Db GAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9709  
Qy AACGCTCCGAGCAGAGTGTATGCGGGGCGCGCGGCAACCCGTCATGCGATCGCGCG 2160  
Db AACGCTCCGAGCAGAGTGTATGCGGGGCGCGCGGCAACCCGTCATGCGATCGCGCG 9769  
Qy AACGCTCCGAGCAGAGTGTATGCGGGGCGCGCGGCAACCCGTCATGCGATCGCGCG 9769  
Db AACGCTCCGAGCAGAGTGTATGCGGGGCGCGCGGCAACCCGTCATGCGATCGCGCG 9769  
Qy GCGATGAGCG 2220  
Db GCGATGAGCG 9829  
Qy CGGCTCATGAGCG 2280  
Db CGGCTCATGAGCG 9889  
Qy CGGCTCATGAGCG 2340  
Db CGGCTCATGAGCG 9949  
Qy CGGCTCATGAGCG 9949  
Db CGGCTCATGAGCG 9949  
Qy CGGCTCATGAGCG 2400  
Db CGGCTCATGAGCG 10009  
Qy CGGCTCATGAGCG 2460  
Db CGGCTCATGAGCG 10069  
Qy GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 2520  
Db GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 10129  
Qy GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 2580  
Db GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 10189  
Qy GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 2640  
Db GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 10249  
Qy GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 2700  
Db GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 10309  
Qy GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 2760  
Db GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 10369  
Qy GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 2820  
Db GGGGTAAGCTGCTGGGATGAGCGGGACCAACGCGCACTGTCGTCGAAATGCAAGCTGCTC 10389





```
Db 7790 GACCCGATGCCCCGGGAGACGCCCCGATCTCTTCTGAGGAGGATGCC 7849
Qy 241 TGTTCGACGCGCTCTTCTTGCGCATCTCGCTCGAGGCGCTGCGGATGACCTTGA 300
Db 7850 TGTTCGACGCGCTCTTCTTGCGCATCTCGCTCGAGGCGCTGCGGATGACCTTGA 7909
Qy 301 CATGCACTCTTGAGAGTGTGCTGAGAGCGCTGAGAAAGCCGCGCATGCTCATCG 360
Db 7910 CATGCACTCTTGAGAGTGTGCTGAGAGCGCTGAGAAAGCCGCGCATGCTCATCG 7969
Qy 361 GCGCTCGTGGTACGGAACGGGAGTTTCAATCGGATCGGCCCCGTAATGAGGCC 420
Db 7970 GCGCTCGTGGTACGGAACGGGAGTTTCAATCGGATCGGCCCCGTAATGAGGCC 8029
Qy 421 GCGCTGCGCAAGGAGCGGCGTCCGAGAGTCGCTCATGCGGCGCTGAGGAGCATG 480
Db 8030 GCGCTGCGCAAGGAGCGGCGTCCGAGAGTCGCTCATGCGGCGCTGAGGAGCATG 8089
Qy 481 CCCAGCGTGGAGGCGGCGGCAATCTGATGCCCCCTCGGCTGCGAGGCGCTGATCGG 540
Db 8090 CCCAGCGTGGAGGCGGCGGCAATCTGATGCCCCCTCGGCTGCGAGGCGCTGATCGG 8149
Qy 541 GTGATACGCGCTTATCTCTCTCGTGTGCGCTTCAATCTGCGCTGTCAGAGCTTGGC 600
Db 8150 GTGATACGCGCTTATCTCTCTCGTGTGCGCTTCAATCTGCGCTGTCAGAGCTTGGC 8209
Qy 601 TCGGGGGAATGCTCAAGGCGCTGCGTGTGAGGATGCGTGAATGTCGCGAGCAAC 660
Db 8210 TCGGGGGAATGCTCAAGGCGCTGCGTGTGAGGATGCGTGAATGTCGCGAGCAAC 8269
Qy 661 CTCTGTGTCTCTGAGAACCGGCGCGTGGCCAGAGGAGCGTGTGAGCAATTTTCG 720
Db 8270 CTCTGTGTCTCTGAGAACCGGCGCGTGGCCAGAGGAGCGTGTGAGCAATTTTCG 8329
Qy 721 GCGGAGCGCATGCGTTTGAAGAGGCGAAGGATGCGCTGCTGATGCTTCAAGCGCTC 780
Db 8330 GCGGAGCGCATGCGTTTGAAGAGGCGAAGGATGCGCTGCTGATGCTTCAAGCGCTC 8389
Qy 781 AGTGAAGCGCGGCGGAGCGGAGTTCGATTTGGCGGATTCAGAGATCCGCGATCAT 840
Db 8390 AGTGAAGCGCGGCGGAGCGGAGTTCGATTTGGCGGATTCAGAGATCCGCGATCAT 8449
Qy 841 CACGACGCTGCGAGCGGCTCTGACCTGTGCGGAACGGAGCTCCAGAAATGCTGTG 900
Db 8450 CACGACGCTGCGAGCGGCTCTGACCTGTGCGGAACGGAGCTCCAGAAATGCTGTG 8509
Qy 901 AAACGGGCGCTGCGGAGCGAGGCTGCGCGCTTCTGAGGATTCGAGGCAAC 960
Db 8510 AAACGGGCGCTGCGGAGCGAGGCTGCGCGCTTCTGAGGATTCGAGGCAAC 8569
Qy 961 GGGACGGGCGAGGCGTGTGAGACCCCATTCGAATCCAAAGCTGGAAGCGGTATACGC 1020
Db 8570 GGGACGGGCGAGGCGTGTGAGACCCCATTCGAATCCAAAGCTGGAAGCGGTATACGC 8629
Qy 1021 CTCGGGCGAGATGTGCGACCGCTGCTGATCGGCTGCTGAGAACCACTTGACAT 1080
Db 8630 CTCGGGCGAGATGTGCGACCGCTGCTGATCGGCTGCTGAGAACCACTTGACAT 8689
Qy 1081 CTTGAGTATGCTGCGGAGATCACTGAGCTGTGAAGTCTGTTCTTCAAGCGAG 1140
Db 8690 CTTGAGTATGCTGCGGAGATCACTGAGCTGTGTGAAGTCTGTTCTTCAAGCGAG 8749
Qy 1141 CAGATTCCTGCGACCTTCAAGCGCGAGCGCTGAACCCCGGAGTTCATGAGGATCTT 1200
Db 8750 CAGATTCCTGCGACCTTCAAGCGCGAGCGCTGAACCCCGGAGTTCATGAGGATCTT 8809
Qy 1201 GCGGTCGACCTTCAAGCGCGCGCGGACACCTGTGCGGACTGGAATACCGCGGAGCG 1260
Db 8810 GCGGTCGACCTTCAAGCGCGCGCGGACACCTGTGCGGACTGGAATACCGCGGAGCG 8869
Qy 1261 GGGGTGACCTGTTGCGCATGAGCGGAGCAACCGCGCATGCTGTGGAAGAGCGCGG 1320
Db 8870 GGGGTGACCTGTTGCGCATGAGCGGAGCAACCGCGCATGCTGTGGAAGAGCGCGG 8929
Qy 1321 GCGGCGAGTGCACACCGCGCGCGCGGAGCGAGCCGCGAGAGCTGTGCTGTGCGCA 1380
Db 8930 GCGGCGAGTGCACACCGCGCGCGCGGAGCGAGCCGCGAGAGCTGTGCTGTGCGCA 8989
Qy 1381 AGGACCGCTGACCTGAGATGCAAGGCGCGCGCTGCGGACCATCTGAGACCTTAC 1440
Db 8990 AGGACCGCTGACCTGAGATGCAAGGCGCGCGCGCTGCGGACCATCTGAGACCTTAC 9049
Qy 1441 CTTTCCGAGTGTCTGCGGCGATGAGGCTTCAATGCTGCGAGCAAGACCGCGCATAGAG 1500
Db 9050 CTTTCCGAGTGTCTGCGGCGATGAGGCTTCAATGCTGCGAGCAAGACCGCGCATAGAG 9109
Qy 1501 CACCGGCTCGCGGAGGCGGAGCGTCAAGGAGGAGGCTGCGGAGCGCTTGAAGCTTGG 1560
Db 9110 CACCGGCTCGCGGAGGCGGAGCGTCAAGGAGGAGGCTGCGGAGCGCTTGAAGCTTGG 9169
Qy 1561 GCGGAGGAGACAGAGTGCAGCGCGGAGTGCAGATTCGCGCATTCACGCGGCAAG 1620
Db 9170 GCGGAGGAGACAGAGTGCAGCGCGGAGTGCAGATTCGCGCATTCACGCGGCAAG 9229
Qy 1621 CTGCGCTTCTCTTACCGGACAGAGGCGCGGAGCGTGGGATAGGCGCTGTGATC 1680
Db 9230 CTGCGCTTCTCTTACCGGACAGAGGCGCGGAGCGTGGGATAGGCGCTGTGATC 9289
Qy 1681 GATGTATGTCGCGGCTTCCGAGAGCGCTTCACTGTGCGGAGGCTTTCACACAGAG 1740
Db 9290 GATGTATGTCGCGGCTTCCGAGAGCGCTTCACTGTGCGGAGGCTTTCACACAGAG 9349
Qy 1741 CTGACCGGCGCTTCCGAGAGTATGAGGCGGACCGGACGCGGAGCGCGCTG 1800
Db 9350 CTGACCGGCGCTTCCGAGAGTATGAGGCGGACCGGACGCGGAGCGCGCTG 9409
Qy 1801 CTGACCGAGACAGCTTCAACCGGCGCGCTGCTTCACTTGAATATGCGCTGCGCG 1860
Db 9410 CTGACCGAGACAGCTTCAACCGGCGCGCTGCTTCACTTGAATATGCGCTGCGCG 9469
Qy 1861 CTGTGCGGCTGCTGAGGCTGAGACCGGAGTGTGCTGCGGCGATGAGATCGGTGAGCTG 1920
Db 9470 CTGTGCGGCTGCTGAGGCTGAGACCGGAGTGTGCTGCGGCGATGAGATCGGTGAGCTG 9529
Qy 1921 GTGCGCTCTGCGTGGCGGCGTGTCTGCTTGAAGACGCGGATTCCTGTGCTGCG 1980
Db 9530 GTGCGCTCTGCGTGGCGGCGTGTCTGCTTGAAGACGCGGATTCCTGTGCTGCG 9589
Qy 1981 GCGGCGCGCTGATGACAGGCGCTGCGCGCGCGGCGGATGCTGATGAGCGCGCG 2040
Db 9590 GCGGCGCGCTGATGACAGGCGCTGCGCGCGCGGCGGATGCTGATGAGCGCGCG 9649
Qy 2041 GAGGCGATGAGCTGCTGCGGAGGCGCGGACGCGATGCTGATGCGCGGATC 2100
Db 9650 GAGGCGATGAGCTGCTGCGGAGGCGCGGACGCGATGCTGATGCGCGGATC 9709
Qy 2101 AACGCTCCGAGCAGGATGATTCGCGGAGCGCGGAGCAACCGGTGATGCGATGCGCGG 2160
Db 9710 AACGCTCCGAGCAGGATGATTCGCGGAGCGCGGAGCAACCGGTGATGCGATGCGCGG 9769
Qy 2161 GCGATGCGCGGCGCGGAGGCGGAGCAACAGGCGCTTCAAGTCTGATGCTTCACTCA 2220
Db 9770 GCGATGCGCGGCGCGGAGGCGGAGCAACAGGCGCTTCAAGTCTGATGCTTCACTCA 9829
Qy 2221 CCGCTCATGAGCGCGGAGTGTGAGAGGCTTGGAGGCTGTGCGGAGTGTGAGCTTACCGG 2280
Db 9830 CCGCTCATGAGCGCGGAGTGTGAGAGGCTTGGAGGCTGTGCGGAGTGTGAGCTTACCGG 9889
Qy 2281 CCGCGCTGATGCTGCTGATCAGCAATCTGAGCGGAGAGCTTTCACAGAGAGTGAAGC 2340
Db 9890 CCGCGCTGATGCTGCTGATCAGCAATCTGAGCGGAGAGCTTTCACAGAGAGTGAAGC 9949
Qy 2341 TCGCGGCGCTATTTGAGTGCAGCGCGCGGAGAGTGTGCTTTCGCGGATGAGTGAAG 2400
Db 9950 TCGCGGCGCTATTTGAGTGCAGCGCGCGGAGAGTGTGCTTTCGCGGATGAGTGAAG 10009
```

QY	2401	GGCGTCGACGCGCGCGCGTGGCGGGACCTTGTCGTGAGCGTGCCTGAAATTCAGCGCTGC	2460
Db	10010	GGCGTCGACGCGCGCGCGTGGCGGGACCTTGTCGTGAGCGTGCCTGAAATTCAGCGCTGC	10069
QY	2461	GGCGTCGTGCTCGTGGCTTCATGCGCGAGCGCCCGGCGCGCTGCTGCATGTCGCGCT	2520
Db	10070	GGCGTCGTGCTCGTGGCTTCATGCGCGAGCGCCCGGCGCGCTGCTGCATGTCGCGCT	10129
QY	2521	GGGGGTACACGACGCGGGGACCGGTCTGAGAGCGCTGCGGGGCTCTGGGCGCGTGGG	2580
Db	10130	GGGGGTACACGACGCGGGGACCGGTCTGAGAGCGCTGCGGGGCTCTGGGCGCGTGGG	10189
QY	2581	CTGGTCTCTTGAGCGCGCTCTTTCCTCTCAAGGGGGGGCGGGTGCCTGCTCCACGTAC	2640
Db	10190	CTGGTCTCTTGAGCGCGCTCTTTCCTCTCAAGGGGGGGCGGGTGCCTGCTCCACGTAC	10249
QY	2641	CCTTGGCAGCGCGACCGCTACTGATGACACGAAACGACGACGCGCGCTGTGGCGAC	2700
Db	10250	CCTTGGCAGCGCGACCGCTACTGATGACACGAAACGACGACGCGCGCTGTGGCGAC	10309
QY	2701	CGCGGTGCTCGGAGAGGGGTCAAGCAGAGGTGAGGAGGGGGCGCGTGCAGCGCGAC	2760
Db	10310	CGCGGTGCTCGGAGAGGGGTCAAGCAGAGGTGAGGAGGGGGCGCGTGCAGCGCGAC	10369
QY	2761	GACCGCGCGACGCTCGGCTTCAGCATGCGCCCGCGAGCGGACGCGCGGAGAAAGTC	2820
Db	10370	GACCGCGCGACGCTCGGCTTCAGCATGCGCCCGCGAGCGGAGCGCGGAGAAAGTC	10429
QY	2821	GAGGCGCGCGCGACCGTCCGCTTCGCGCTGAGATGCAATAGCCAGCGCGTGTGATCAC	2880
Db	10430	GAGGCGCGCGCGACCGTCCGCTTCGCGCTGAGATGCAATAGCCAGCGCGTGTGATCAC	10489
QY	2881	CTCGTCTCTGGGTACAGAGCGCGCGCCCTGTGTCTGGGCGAGGTGCAGATCGCGCT	2940
Db	10490	CTCGTCTCTGGGTACAGAGCGCGCGCCCTGTGTCTGGGCGAGGTGCAGATCGCGCT	10549
QY	2941	GACCGCGCGGGGCTCACGCTTCAGATGATCAGTCCGCTGGGCAATGTCGCCGACGAC	3000
Db	10550	GACCGCGCGGGGCTCACGCTTCAGATGATCAGTCCGCTGGGCAATGTCGCCGACGAC	10609
QY	3001	CTGCGCGGAAAGCCCAACCCCTCGCTGCTGTGAGAGCGAGTGCGCCGCGACATGCT	3060
Db	10610	CTGCGCGGAAAGCCCAACCCCTCGCTGCTGTGAGAGCGAGTGCGCCGCGGACATGCT	10669
QY	3061	GCCGTGGGCGAGGGCGGTGAAGAGGCTCTGTGTGGGCCAACCCGTCATCGCCCTTTGGCG	3120
Db	10670	GCCGTGGGCGAGGGCGGTGAAGAGGCTCTGTGTGGGCCAACCCGTCATCGCCCTTTGGCG	10729
QY	3121	GGACCGTATTGTATCCACAGTCAACCACTGCGCTGTGTGTGTGCTCGACTCAGGCG	3180
Db	10730	GGACCGTATTGTATCCACAGTCAACCACTGCGCTGTGTGTGTGCTCGACTCAGGCG	10789
QY	3181	CTTTCGCGCATGAGCGCGCGCCCATCCCGTGCCTGACTTGACGCGCATGTACGCGCTC	3240
Db	10790	CTTTCGCGCATGAGCGCGCGCCCATCCCGTGCCTGACTTGACGCGCATGTACGCGCTC	10849
QY	3241	GACGCAATAGCGCGGCTTTCAGCGCGGGGAGCGGGTGTATTCATCGCGCGACCGCGGG	3300
Db	10850	GACGCAATAGCGCGGCTTTCAGCGCGGGGAGCGGGTGTATTCATCGCGCGACCGCGGG	10909
QY	3301	GTCGGTCTCGCGCGGTGACAGTGGGCCACAGCAGTGGAGCCGAGGTTCATGCGACGCGC	3360
Db	10910	GTCGGTCTCGCGCGGTGACAGTGGGCCACAGCAGTGGAGCCGAGGTTCATGCGACGCGC	10969
QY	3361	GGCAGCGCGGAGAAACGCGCTTACTGTGAAGTCTGTGGGCGTGGGTATGTAGACCATTC	3420
Db	10970	GGCAGCGCGGAGAAACGCGCTTACTGTGAAGTCTGTGGGCGTGGGTATGTAGACCATTC	11029
QY	3421	CGCTTCGACCGGTTCTGTGCGCGAGAGTGCAGCGCGCTGTGACCGGCGCGAGGAGTACGCTC	3480
Db	11030	CGCTTCGACCGGTTCTGTGCGCGAGAGTGCAGCGCGCGTGTGACCGGCGCGAGGAGTACGCTC	11089

QY	3481	GTGCTCAACTGGCTCTCGGCGAGCTGATCGACAAAGATTTCATCTCTCGGATCGCAC	3540
Db	11090	GTGCTCAACTGGCTCTCGGCGAGCTGATCGACAAAGATTTCATCTCTCGGATCGCAC	11149
QY	3541	GGCCGGTTTGTGAGAGCTCGGCAAGCCGACCTGTTCAGCGGATTAACCAAGCTCGGCTGCGG	3600
Db	11150	GGCCGGTTTGTGAGAGCTCGGCAAGCCGACCTGTTCAGCGGATTAACCAAGCTCGGCTGCGG	11209
QY	3601	CCGTTCTCGGCAATCTCTCTTCTGCGCTGAGATCTCCGGGGGATGATGCTTCGAGCCG	3660
Db	11210	CCGTTCTCGGCAATCTCTCTTCTGCGCTGAGATCTCCGGGGGATGATGCTTCGAGCCG	11269
QY	3661	CCGGCGCGGGTTCGGTGCGCTCTTGAGAGAGCTCTGAGGCTGATCGCGCAGGCGGTTTC	3720
Db	11270	CCGGCGCGGGTTCGGTGCGCTCTTGAGAGAGCTCTGAGGCTGATCGCGCAGGCGGTTTC	11329
QY	3721	ACCCCTCCCCCATTCGAGACGCTCCCGATCGCCGTCGTGCGCGATGCGTTTCGGAGCATG	3780
Db	11330	ACCCCTCCCCCATTCGAGAGCTCCCGATCGCCGTCGTGCGCGATGCGTTTCGGAGCATG	11389
QY	3781	GCGCAGGCGCAGCATCTTGGGAAGCTGTACTCAACGCTGAGTGACCCGAGGTTCCAGATC	3840
Db	11390	GCGCAGGCGCAGCATCTTGGGAAGCTGTACTCAACGCTGAGTGACCCGAGGTTCCAGATC	11449
QY	3841	CGATTTCACACCAACGACGAGCGCGCGCCGCTCCACCGGGGATTCGGGAACTGCTCGACAG	3900
Db	11450	CGATTTCACACCAACGACGAGCGCGCGCCGCTCCACCGGGGATTCGGGAACTGCTCGACAG	11509
QY	3901	CTCGCTCACTCGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCTTCGTACGCGAG	3960
Db	11510	CTCGCTCACTCGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCTTCGTACGCGAG	11569
QY	3961	GTCTCGCAGGTGCTGCGCACGCCCGAATCAAGGTGCGCGAGGCGCTGTTCAACCCGC	4020
Db	11570	GTCTCGCAGGTGCTGCGCACGCCCGAATCAAGGTGCGCGAGGCGCTGTTCAACCCGC	11629
QY	4021	CTCGGCAATGCAATCTGCTCAATGAGCGGTGAGAGTCGCGCAATCGTATCGAGGCGAGCTCAAG	4080
Db	11630	CTCGGCAATGCAATCTGCTCAATGAGCGGTGAGAGTCGCGCAATCGTATCGAGGCGAGCTCAAG	11689
QY	4081	CTGAAGCTGTGACGACAGGTTCTCTGTCACAGTCCGCCAATATGCGCTGTTGGCCCAAAAC	4140
Db	11690	CTGAAGCTGTGACGACAGGTTCTCTGTCACAGTCCGCCAATATGCGCTGTTGGCCCAAAAC	11749
QY	4141	CTGTTGAGATCTCTCGGCACAGCTTCTCTTTGAGCGGGGTGCGCGGAGAACTTACCG	4200
Db	11750	CTGTTGAGATCTCTCGGCACAGCTTCTCTTTGAGCGGGGTGCGCGGAGAACTTACCG	11809
QY	4201	GCAGGCGTGCMAAACGACTTCTGCTCATCGGGGCGAGTCAAGACTGGGAATATCATTTGCC	4260
Db	11810	GCAGGCGTGCMAAACGACTTCTGCTCATCGGGGCGAGTCAAGACTGGGAATATCATTTGCC	11869
QY	4261	CTATGCA 4266	
Db	11870	CTATGCA 11875	
RESULT 8			
US-09-443-501A-2			
; Sequence 2, Application US/09443501A			
; Patent No. 6103342			
; GENERAL INFORMATION:			
; APPLICANT: Kosan Biosciences, Inc.			
; APPLICANT: Julien, Bryan			
; APPLICANT: Katz, Leonard			
; APPLICANT: Khosla, Chaitan			
; APPLICANT: Tang, Li			
; APPLICANT: Ziermann, Rainer			
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing			
; FILE REFERENCE: 30062-20031.00			
; CURRENT FILING DATE: US/09/443,501A			
; CURRENT FILING DATE: 1999-11-19			

PRIOR APPLICATION NUMBER: US 60/130,560  
PRIOR FILING DATE: 1999-04-22  
PRIOR APPLICATION NUMBER: US 60/122,620  
PRIOR FILING DATE: 1999-03-03  
PRIOR APPLICATION NUMBER: US 60/119,386  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: US 60/109,401  
PRIOR FILING DATE: 1998-11-20  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 2  
LENGTH: 71989  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic construct  
US-09-443-501A-2

Query Match 98.2%; Score 4187.6; DB 4; Length 71989;  
Best Local Similarity 98.9%; Pred. No. 0;  
Matches 4217; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 GTGGGCGATCTCCCATCGAGCGCCGACGCCGAAGATCCGATTCCGATGCTCGAGCGAGT 60  
DB 1998 GTGGGCGATCTCCCATCGAGCGCCGACGCCGAAGATCCGATTCCGATGCTCGAGCGAGC 2057  
QY 61 TGCCGCTGCGCCGCGTGGCGGTGATCGATCGAGCGGGTTCTGAGCGCTCTCGAGGGCTCG 120  
DB 2058 TGCCGCTGCGCCGCGTGGCGGTGATCGATCGAGCGGGTTCTGAGCGCTCTCGAGGGCTCG 2117  
QY 121 CGCGACACCGTGGCGAGTCCCGCGCAACGCTGGGATGACAGCGGTGTTGATCC 180  
DB 2118 CGCGACACCGTGGCGAGTCCCGCGCAACGCTGGGATGACAGCGGTGTTGATCC 2177  
QY 181 GACCCCGATGCCCGGGGAAAGCGCCGTTACGCGCGCATCTTTCTGAGCGAGTACG 240  
DB 2178 GACCTCGATGCCCCCGGGGAAAGCGCCGTTACGCGCGCATCTTTCTGAGCGAGTACG 2237  
QY 241 TGCTTGAAGCGCTCTCTTCTGGGCACTCTGCGCGAAGGCGTGGATGAGACCTGCA 300  
DB 2238 TGCTTGAAGCGCTCTCTTCTGGGCACTCTGCGCGAAGGCGTGGATGAGACCTGCA 2297  
QY 301 CATCGACTTCTGAGAGGTGCTGAGAGCGCTGAGAGACGCGCGATCGCTCCATCG 360  
DB 2298 CATCGACTTCTGAGAGGTGCTGAGAGCGCTGAGAGACGCGCGATCGCTCCATCG 2357  
QY 361 GCGCTCTGCTGAGGAAACGGAGTGTTCATCGGGATGCGCCCGTCCGAATATGAGGCC 420  
DB 2358 GCGCTCTGCTGAGGAAACGGAGTGTTCATCGGGATGCGCCCGTCCGAATATGAGGCC 2417  
QY 421 GCGCTGCGGCAAGCAGCGCGCTCCGAGAGATCGACGCTCATGGGCGGTGGGAGCGATG 480  
DB 2418 GCGCTGCGGCAAGCAGCGCGCTCCGAGAGATCGACGCTCATGGGCGGTGGGAGCGATG 2477  
QY 481 CCCAGCGTGGAGCGGCGGCAATCTGTATGCCCTCGAGGCTCGAGAGCGCTGTCTCGG 540  
DB 2478 CCCAGCGTGGAGCGGCGGCAATCTGTATGCCCTCGAGGCTCGAGAGCGCTGTCTCGG 2537  
QY 541 GTGATAGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
DB 2538 GTGATAGCGGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2597  
QY 601 TCCGGGGAATGCTCAACGCGCGCTGGCTGGTGGGATATGCTGATGCTGCTGCTGCTGCTG 660  
DB 2598 TCCGGGGAATGCTCAACGCGCGCTGGCTGGTGGGATATGCTGATGCTGCTGCTGCTGCTG 2657  
QY 661 CTCGTGTGGCTCTCGAAGACCGCGCGCTGGCGAGGACGCTGCTGCAAGGCAATTTTCG 720  
DB 2658 CTCGTGTGGCTCTCGAAGACCGCGCGCTGGCGAGGACGCTGCTGCAAGGCAATTTTCG 2717  
QY 721 GCGGAGCGCGATGGGTTGCGAGCGAGCGAGGAGGTGCGCTGTGTGCTCTCAAGCGGCTC 780  
DB 2718 GCGGAGCGCGATGGGTTGCGAGCGAGCGAGGAGGTGCGCTGTGTGCTCTCAAGCGGCTC 2777

QY 781 AGTGAAGCCCGCGCGAGCGGATCGAATATTTGGCGTGAATTCAGAGATCCGCGATCAAT 840  
DB 2778 AGTGAAGCCCGCGCGAGCGGATCGAATATTTGGCGTGAATTCAGAGATCCGCGATCAAT 2837  
QY 841 CACGACGCTGGAGCAGCGGCTGTGAACCGTCCGAAACGGAGCTCCCAAGAAATGCTGCTG 900  
DB 2838 CACGACGCTGGAGCAGCGGCTGTGAACCGTCCGAAACGGAGCTCCCAAGAAATGCTGCTG 2897  
QY 901 AAACGGGCGCTGGGGAAGCGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
DB 2898 AAACGGGCGCTGGGGAAGCGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2957  
QY 961 GGCACGGGCGAGCGCTTGTGACCCCAATCGAAATTCAGGCTTGAATGCGGTATACGCG 1020  
DB 2958 GGCACGGGCGAGCGCTTGTGACCCCAATCGAAATTCAGGCTTGAATGCGGTATACGCG 3017  
QY 1021 CTCGGGCGAGATGTCGACCGCGCTGCTGATTCGGGCTGGTGAAGACCAACTTGGCCAT 1080  
DB 3018 CTCGGGCGAGATGTCGACCGCGCTGCTGATTCGGGCTGGTGAAGACCAACTTGGCCAT 3077  
QY 1081 CCGATGATGCGTGGGGAATCACTGGGCTGCTGAAGGTGCTGCTGCTGCTGCTGCTGCTG 1140  
DB 3078 CCGATGATGCGTGGGGAATCACTGGGCTGCTGAAGGTGCTGCTGCTGCTGCTGCTGCTG 3137  
QY 1141 CAGATTCTGCGCACTTCACGCGAGCGGCTGAACCCCGGATCTCAATGGGATGATCTT 1200  
DB 3138 CAGATTCTGCGCACTTCACGCGAGCGGCTGAACCCCGGATCTCAATGGGATGATCTT 3197  
QY 1201 CGGCTGACCTGACCGCGCGCGCGGACACCTGTCGCGGACTGGAATACCCCGGACGCGG 1260  
DB 3198 CGGCTGACCTGACCGCGCGCGCGGACACCTGTCGCGGACTGGAATACCCCGGACGCGG 3257  
QY 1261 GGGGTGACCTGCTTGGGATAGCGGGAACAAACGCGACGTGTGTGGAAGAGCGCGCG 1320  
DB 3258 GGGGTGACCTGCTTGGGATAGCGGGAACAAACGCGACGTGTGTGGAAGAGCGCGCG 3317  
QY 1321 GCGGCGAGTGCACACCGCGCGCGCGGAGGACCGGAGAGCTGCTGCTGCTGCGCA 1380  
DB 3318 GCGGCGAGTGCACACCGCGCGCGCGGAGGACCGGAGAGCTGCTGCTGCTGCGCA 3377  
QY 1381 AGGACCGGTGACCGCTGGATGACAGCGCGCGCGCTGCGGACCATCTGAGACTTAC 1440  
DB 3378 AGGACCGGTGACCGCTGGATGACAGCGCGCGCGCTGCGGACCATCTGAGACTTAC 3437  
QY 1441 CCTTGGAGTGTCTGGGAGATGCTGCGCTTCACTGTGGGAGAGACCGCGGACGATGAG 1500  
DB 3438 CCTTGGAGTGTCTGGGAGATGCTGCGCTTCACTGTGGGAGAGACCGCGGACGATGAG 3497  
QY 1501 CACCGGCTCGGAGTGGCGGACGTCGAGGAGGAGGCTGCGGAGAGCTTGAACGCTGCG 1560  
DB 3498 CACCGGCTCGGAGTGGCGGACGTCGAGGAGGAGGCTGCGGAGAGCTTGAACGCTGCG 3557  
QY 1561 GCGCAGGAGACAGCGTCCCGCTGGTGGCGAGATTCGCGATTCCTCAACGCGGCAAG 1620  
DB 3558 GCGCAGGAGACAGCGTCCCGCTGGTGGCGAGATTCGCGATTCCTCAACGCGGCAAG 3617  
QY 1621 CTCGCTTCTTCTTCAACGAGGAGGCGGAGAGCTGGGCAATGGGCTGGGCTGTAC 1680  
DB 3618 CTCGCTTCTTCTTCAACGAGGAGGCGGAGAGCTGGGCAATGGGCTGGGCTGTAC 3677  
QY 1681 GATGATGATGCGGCTTCCGAGAGGCTTCAACCTGTCGTCGTCGTCGTCGTCGTCGTCG 1740  
DB 3678 GATGATGATGCGGCTTCCGAGAGGCTTCAACCTGTCGTCGTCGTCGTCGTCGTCGTCG 3737  
QY 1741 CTCGACCGGCGCTTCCGAGAGGATGTCGAGCGGACCGGACAGGCTGACCGCGCTG 1800  
DB 3738 CTCGACCGGCGCTTCCGAGAGGATGTCGAGCGGACCGGACAGGCTGACCGCGCTG 3797  
QY 1801 CTCGACCGAGACGCTTCAACCAAGCGGCGCTGTCCTTCACTTCACTTCACTTCACTTCA 1860  
DB 3798 CTCGACCGAGACGCTTCAACCAAGCGGCGCTGTCCTTCACTTCACTTCACTTCACTTCA 3857

QY 1861 CTGTGGCGGTCGTCGGGGTGTGAGAGCGGAGTTGTGTCCCGGCCCATGATCGGTGAGCTG 1920  
DB |||||  
DB 3858 CTGTGGCGGTCGTCGGGGTGTGAGAGCGGAGTTGTGTGTGCTGCTGCAATGATCGGTGAGCTG 3917  
QY 1921 GTGTGTCGTCGGGTGCGGGGCGGTGTCTGTGCTGAGAGACGCGGTGTTCTGTGTGCTGCG 1980  
DB |||||  
DB 3918 GTGTGTCGTCGGGTGCGGGGCGGTGTCTGTGCTGAGAGACGCGGTGTTCTGTGTGCTGCG 3977  
QY 1981 CCGGGGCGCTGTGATGCAAGCGCTGTCCGCGCGGGCGGATGCTGTGATCGAGCGCGG 2040  
DB |||||  
DB 3978 CCGGGGCGCTGTGATGCAAGCGCTGTCCGCGCGGGCGGATGCTGTGATCGAGCGCGG 4037  
QY 2041 GAGGCGGATGTGGCTGTCCGCGGTGCGCGGACGCAAGCGGTGCTGTGATCGCGCGGCTC 2100  
DB |||||  
DB 4038 GAGGCGGATGTGGCTGTCCGCGGTGCGCGGACGCAAGCGGTGCTGTGATCGCGCGGCTC 4097  
QY 2101 AACGCTCCGAGACGAGGTGTGATCGCGGGGCGCGGCAACCGGTGATCGATCGCGGG 2160  
DB 4098 AACGCTCCGAGACGAGGTGTGATCGCGGGGCGCGGCAACCGGTGATCGATCGCGGG 4157  
QY 2161 GCGATGCGCGCGCGCGGGCGGAGACCAAGCGCTGTGATCGATCGCTGTGATCGCTGCACTCA 2220  
DB 4158 GCGATGCGCGCGCGCGGGCGGAGACCAAGCGCTGTGATCGATCGCTGTGATCGCTGCACTCA 4217  
QY 2221 CCGCTCATGAGCCCGATGCTGTGAGAGCGGTTCCGGGCGGTGTGCGGAGTCCGATGAGCTACCG 2280  
DB 4218 CCGCTCATGAGCCCGATGCTGTGAGAGCGGTTCCGGGCGGTGTGCGGAGTCCGATGAGCTACCG 4277  
QY 2281 CCGCGCTCATGCTGTGTGTGATGCAATGTGAGCGGAGAGCTGTGACAGACGAGGTGAGC 2340  
DB 4278 CCGCGCTCATGCTGTGTGTGATGCAATGTGAGCGGAGAGCTGTGACAGACGAGGTGAGC 4337  
QY 2341 TCGCGCGGCTTATTTGGGTGCGGCAACCGCGGAGAGTGTGCTGTGCGGATGAGTGAAG 2400  
DB 4338 TCGCGCGGCTTATTTGGGTGCGGCAACCGCGGAGAGTGTGCTGTGCGGATGAGTGAAG 4397  
QY 2401 GCGCTGCAAGCGGCGGTGCGGGGCACTTGTGTGAGTGTGCTGTGCAATGTGAGCTGCTC 2460  
DB 4398 GCGCTGCAAGCGGCGGTGCGGGGCACTTGTGTGAGTGTGCTGTGCAATGTGAGCTGCTC 4457  
QY 2461 GCGCTGTGTGCTGTGCTGTGATGCAATGTGAGCGGCGGCGCTGTGCAATGTGAGCTGCTC 2520  
DB 4458 GCGCTGTGTGCTGTGCTGTGATGCAATGTGAGCGGCGGCGCTGTGCAATGTGAGCTGCTC 4517  
QY 2521 GCGCGTGTGATGAGCGGCGGACCGTGTGTGAGCGGCTGTGCGGCGCTGTGCGGATGAGTGAAG 2580  
DB 4518 GCGCGTGTGATGAGCGGCGGACCGTGTGTGAGCGGCTGTGCGGCGCTGTGCGGATGAGTGAAG 4577  
QY 2581 CTGTGTCTCTGTGGGCGGCGCTTCTTCCCTCAGGGGAGGCGGCGGCTGTGCGGCGGCTGCTC 2640  
DB 4578 CTGTGTCTCTGTGGGCGGCGCTTCTTCCCTCAGGGGAGGCGGCGGCTGTGCGGCGGCTGCTC 4637  
QY 2641 CTTTGGCAGCGGCGGAGCTTGTGATGCAACGAAAGCGGACGACGCGGCGGCTGTGCGGAC 2700  
DB 4638 CTTTGGCAGCGGCGGAGCTTGTGATGCAACGAAAGCGGACGACGCGGCGGCTGTGCGGAC 4697  
QY 2701 CCGCGTGTGCTGCGGAGCGGCTGTGAGCGGAGTGTGAGAGAGGCGGCGGCTGTGCGGCGG 2760  
DB 4698 CCGCGTGTGCTGCGGAGCGGCTGTGAGCGGAGTGTGAGAGAGGCGGCGGCTGTGCGGCGG 4757  
QY 2761 GACCGGCGGAGCGGCTGTGAGCGGCTGTGAGCGGAGTGTGAGAGAGGCGGCGGCTGTGCGG 2820  
DB 4758 GACCGGCGGAGCGGCTGTGAGCGGCTGTGAGCGGAGTGTGAGAGAGGCGGCGGCTGTGCGG 4817  
QY 2821 GAGGCGGCGGAGCGGCTGTGAGCGGCTGTGAGCGGAGTGTGAGAGAGGCGGCGGCTGTGCGG 2880  
DB 4818 GAGGCGGCGGAGCGGCTGTGAGCGGCTGTGAGCGGAGTGTGAGAGAGGCGGCGGCTGTGCGG 4877  
QY 2881 CTGTGTCTTGTGAGTGTGAGCGGCGGCGGCTGTGAGCGGAGTGTGAGAGTGTGAGAGTGTGAG 2940  
DB 4878 CTGTGTCTTGTGAGTGTGAGCGGCGGCGGCTGTGAGCGGAGTGTGAGAGTGTGAGAGTGTGAG 4937  
QY 2941 GACGCGGCGGCGGCTGTGAGTGTGAGCGGCTGTGAGCGGAGTGTGAGAGTGTGAGAGTGTGAG 3000

DB 4938 GACGCGGCGGCGGCTGTGAGTGTGAGCGGCTGTGAGCGGAGTGTGAGAGTGTGAGAGTGTGAG 4997  
QY 3001 CTGTGCGGAGAAACCCAACTTGTGCTGTGCTGTGAGAGCGAGTGTGCGGCGGATGCTC 3060  
DB 4998 CTGTGCGGAGAAACCCAACTTGTGCTGTGCTGTGAGAGCGAGTGTGCGGCGGATGCTC 5057  
QY 3061 GCGGTGTGAGAGGCTGTGAGCGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAG 3120  
DB 5058 GCGGTGTGAGAGGCTGTGAGCGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAG 5117  
QY 3121 GAGAGCTTGTGCTACCAAGTGTGAGCGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAGGCTGTG 3180  
DB 5118 GAGAGCTTGTGCTACCAAGTGTGAGCGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAGGCTGTG 5177  
QY 3181 CTGTGCGGAGTGTGAGAGGCTGTGAGCGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAGGCTGTG 3240  
DB 5178 CTGTGCGGAGTGTGAGAGGCTGTGAGCGGCTGTGAGAGGCTGTGAGAGGCTGTGAGAGGCTGTG 5237  
QY 3241 GACAGAAATGCGGCTTCAAGCGGCGGAGGAGGCTGTGAGTGTGAGCGGCGGCGGCGG 3300  
DB 5238 GACAGAAATGCGGCTTCAAGCGGCGGAGGAGGCTGTGAGTGTGAGCGGCGGCGGCGGCGG 5297  
QY 3301 GTGTGCTGTGCGGCGGAGTGTGAGCGGCGGAGGCTGTGAGTGTGAGCGGCGGCGGCGGCGG 3360  
DB 5298 GTGTGCTGTGCGGCGGAGTGTGAGCGGCGGAGGCTGTGAGTGTGAGCGGCGGCGGCGGCGG 5357  
QY 3361 GCGACGCGCGGAGAAACCGGCTTACTGTGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGG 3420  
DB 5358 GCGACGCGCGGAGAAACCGGCTTACTGTGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGG 5417  
QY 3421 CCGTGTGAGCGGCTTGTGAGCGGCGGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3480  
DB 5418 CCGTGTGAGCGGCTTGTGAGCGGCGGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5477  
QY 3481 GTGTCAACTGTGCTGTGAGGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3540  
DB 5478 GTGTCAACTGTGCTTGTGAGGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5537  
QY 3541 GCGCGGCTTGTGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3600  
DB 5538 GCGCGGCTTGTGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 5597  
QY 3601 CCGTGTGAGCGGCTTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3660  
DB 5598 CCGTGTGAGCGGCTTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 5657  
QY 3661 CCGGCGGAGTGTGAGCGGCTTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3720  
DB 5658 CCGGCGGAGTGTGAGCGGCTTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 5717  
QY 3721 ACCCTTCCCGCATGTGAGCGGCTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3780  
DB 5718 ACCCTTCCCGCATGTGAGCGGCTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 5777  
QY 3781 GCGGAGGCGGAGTGTGAGCGGCTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3840  
DB 5778 GCGGAGGCGGAGTGTGAGCGGCTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 5837  
QY 3841 CGTATTCCAAACCGGAGCGGCGGCGGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAG 3900  
DB 5838 CGTATTCCAAACCGGAGCGGCGGCGGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAG 5897  
QY 3901 CTGTGCTGAGTGTGAGCGGCGGCGGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3960  
DB 5898 CTGTGCTGAGTGTGAGCGGCGGCGGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 5957  
QY 3961 GTGTGAGTGTGAGCGGCGGCGGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4020  
DB 5958 GTGTGAGTGTGAGCGGCGGCGGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 6017  
QY 4021 CTGTGAGTGTGAGCGGCGGCGGAGTGTGAGCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4080

[illegible]

```

RESULT 9
US-09-266-965-19
; Sequence 19, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mlcomycin biosynthetic gene cluster
; FILE REFERENCE: 600, 456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 3978
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-19

```

Query Match	19.1%;	Score 814.2;	DB 4;	Length 3978;
Best Local Similarity	57.6%;	Pred. No. 3.3e-134;		
Matches 1549;	Conservative	0;	Mismatches 1108;	Indels 30; Gaps 4;

Qy	18	CGAGCGCCAGCCGAAAGATCCGATTCCGATCGTCGAGGAGATTGCGTCGTCGCCGTG	77
Db	90	CGAGACCCGCCGCCCGAGCCCGTCGCATTCGTGGGCAATGGCGTGCCTTCCCCGGGGA	149
Qy	78	CGTATCGATCTGAACCGGGTTCTGAGACGCTCTCGAGGGCTCGCGGACACCGTCGAGCG	137
Db	150	CGTGGCATCGCCCGAGACGACTGTGGAGCTGTGGTCGCGGAGGGGCGGGAGCGCGTCA	209
Qy	138	AGTCCCCCGCCGAA--GCTGGGATGCAAGACAGCTGTGATTGATCCCGACCCGATGCC	194
Db	210	GTTCCCGCCCGACCGGGGTGGGAAGTCACGCCGCTTCACGACCCCGAGCCGGGCA	269
Qy	195	GGGGAAGACGCCCTTTAAGCGCGCATCTTTCGAGAGCAAGTAGCCTGCTCGAGCGCTC	254
Db	270	GGGCAAGACGTACGCGCGCCACAGCGCGGCTTCTCAAGAGACGCCCGGATTCGACCCGC	329
Qy	255	CTTCTTCGACATCTCGCCTTCGCGAAAGCGCTGCGGATGACCTTCGACATTCGCTT	314
Db	330	CTTCTTCGACATCACGCGCGCGAGCGCTTCGCGATGAGACCGGACAGGAGCATGATCAT	389
Qy	315	GGAGTGTCTGGAGAGCGCTGAGAAACGCCCGGATGCTCCATCGAGCGTCTGTGGTAC	374

Dp	390	GGAGGTCTCTGGGAGGGGCTTTCGAGCAGAGCGGGGCTCGACGCGACCAACTCTGGGGGCGA	449
Oy	375	GGAAACGGAGATGTTCAATCGGATCGGCGCTCCGATATATAGGCGGCGCTGCCGACAC	434
Dp	450	GGAGGTGGGGCTCTTCTGTCGGCTCCAAACAGCAAGACTACTGATCAACGTGCTCGAGCG	509
Oy	435	GAGCGGCTCCGCAAGATTCAGCGCTCATGGCGGGCTGGGGAGCATGCCACAGTTCGAGC	494
Dp	510	GGGGAGCGTCGCGGAGGGCTTC-----ATGGGAGCCGGCACTTCGCGACAGATCTCTC	563
Oy	495	GGGCGGATCTCTGATGCGCTCCGCGCTGCAGAGGCGCGTGTGTCGCGGATGATACGGCTTA	554
Dp	564	CGGCGCGGTCCCTCAACACTTCGGCTTCGAGGGCGCGGCGGTGTGTCGACACCGCGTCG	623
Oy	555	TTGCTCTCTGCTGGTGGCGCTTCAATCGGCTGTTCAGAGCTTGGCGCTCCGAGGAATGCTC	614
Dp	624	CTCTCTCTGCTGGTTCGCGCTGCACCTGGCGCGGCAAGTCTCTGGGCGAGGGGAGTGTCTC	683
Oy	615	CACGGCCCTGGCTGGTGGGGATATGGCTGATTTGTTCGCGAGCAACCTCGTGTGCTCTC	674
Dp	684	CTTGGCCCTGGCGGGCGCGCGAGAGTGGCCACGCGCAACGCTTTCATTCGATTTAG	743
Oy	675	GAAAGACCCGGGCGCTGGCGAGGAGCGGTGCTGCAAGGCAATTTTCGCGGAGAGCGGATGG	734
Dp	744	CCGCGACGCGGGGCTTGCGCCCCCGACGGCGCTGCAAGTCTTCTTCGCGACCGCGACGG	803
Oy	735	GTTTCGACGAGAGCGAAGGATCGCGCTCTGATGCTCTCAAGCGGCTCAATGAGCCCGGC	794
Dp	804	CACCACTGTGTCGAGGGCGCGCGCTGTGCTCTGCGGCTCTCGAGCGCCCGCG	863
Oy	795	GGACGCGGATTCGATATTGGCGGATTCGAGGATCCGCGATCAATCAACAGCGGTGCAG	854
Dp	864	CTTGGGCTTACCCCTGTGCACGCGGTATCTCCGGGCGAGCGCCGTCAACAGAGCGGGCGAG	923
Oy	855	CAGCGGTCTGACGCTGCAGAACGGGAGCTCCCAAGAAATCGTGTGAAACGGGCGCTGGC	914
Dp	924	CGCGGCTTGAACCGCGCCCAACGGAACGGGCGCAACGGGGTGTATCGGCAAGCACTGGC	983
Oy	915	GGACGAGGCTGCGCGCGCTTTCGATGGATTATGTCCAGGCAACCGGCAACGGGCAACGAC	974
Dp	984	CAAGCGCAACGGTGAACGGCGCAACAGCTGTGACGCGGTGAGGCAACCGGCACCGGCAACC	1043
Oy	975	GCTTGTGTGACCCCATTCGAATTCGAAGCTCTGAATGCGGTATACGGCTTCGGGCGAGATGT	1033
Dp	1044	GCTGGGCGACCCGATTCGAGGCCCGCGGCTCTCGGCACTTAACGGGCGGGCGCGGAGGA	1103
Oy	1035	CGGCACCCGCTGTGATTCGGGTGGGTGAGAACCAACTTGGGCACTCTGATATGCGTC	1099
Dp	1104	GGGAGGCGCGTGTGGCTGGGCTGTGAATTCGAACCTGGGCAACCAAGTCCGCGGCG	1166
Oy	1095	GGGATCATCTGGGCTGCTGAAGTTCGTTCTTCCTTCAGACGGGCGAGTTCTTCGCGCA	1155
Dp	1164	CGGCGCGGGCGGCGTCAATCAAGATGTGTATGGCCATGGGCAACGGGACGCTGCCCGGAC	1222
Oy	1155	CCTTCACGCGGAGGCGCTGAAACCCCGGATCTCATGGGATCTTTCGCTGACCGTCA	1211
Dp	1224	GCTGCACCTTCAGAGGCCACCCCGCGGTGCACTGGTCCGCGGTACGTACGGCTCT	1283
Oy	1215	GCGGCGCCGGAACCGGTGGCTGGACTGGAATACGCGCGGACGGGCGGGGGTGAAGTCTGTT	1277
Dp	1284	GACCGAGGCGCCAGGACTGGCGGAGCAACCGGACACCGCGCGGTGCGGCGGTCTCGTCTT	1343
Oy	1275	CGGATGAGCGGAGCAACACGGCGCATGTGTGCTGGAAGAGCGCGCGGGCGGACCTGAC	1333
Dp	1344	CGGCGTCAAGCGGACCAACGCCCATGTGTATCTTGGAGGGCGCGCGCGCGGAGAGGCAAC	1400
Oy	1335	ACCGCG-----GGCGCGAGCGAACCGGCAAGCTGTGTGTGTCTC	1376
Dp	1404	GGAAGCGCGCGCTGCGGAGCGTCTCTCGAGAGCCCGGGGCGCGGCTGTGGGTGTCTTC	1463
Oy	1377	GGCAAGGACCGGCTCAACCTCTGATTCGACAGAGCGCGCGCGGTGGCGGCACTTGGAGAC	1433
Dp	1464	CGGCGGACGCGGCGCGCTGTCCGAGCGCAAGCGGAGGCTTGGCGGCGCCACTTGAACCGC	1522



QY 1437 CTACCTTTCGAGTGTGTGGGCGATGTGGCTTCACTCTGGGAGCAGACCGCGAGCGCAT 1496  
DB 1524 GCGCCCGCAGCTGGGACCGGCGGACCTGGGCGACCGGCTGGGCGACACCGGCGGCGCTT 1583  
QY 1497 GAGGACCGGCTCGCGGTGGCGGCGAGCTGAGAGGAGGGGCTGGGGGAGCCCTGGAGCG 1556  
DB 1584 CGACACCGGGGCGCGCTGTGTGGCGGGAGCGGTAGAGAACTGTCTGGCGGCTCGCGGCG 1643  
QY 1557 TCGCGGCGAGGAGACAGACGTGCGCGCGGTGGCGGCGAGTATGCCGATTCTTCAAGCGG 1616  
DB 1644 CTGGGCGACCGGAAACCGCGCGCGCGCTGGGTACCGGCGGAGCCCGCGCTCGCGGCG 1703  
QY 1617 CAGAGCTCGGCTTCTTCTTTCACCGGAGAGGGGCGGAGCGCTGGGAGATGGGCGTGGGCT 1676  
DB 1704 CAGGCGCGGCTTCTTCTTTCACCGGAGAGGGCGGAGCGGCGGATGGGCGCGCAACT 1763  
QY 1677 GTACGATGATGGTTCGCGGCTTCGCGAGGCGGTTGACCTGTGGGTGAGGCTGTTCACCA 1736  
DB 1764 GCGGCTTCACAGACCTGTGTGGCGGACCGCTTGGAGAGGTGTGCGCCGAGCTCGACCG 1823  
QY 1737 GAGGCTCGACCGGCGGCTTCGCGAGGTGATGTGGGCGGAAACCGGCGAGCGTGCACCGCG 1796  
DB 1824 GCACCTCGACCGGCGGCTTCGCGAGGTGCTGTTCGCGCGGAGCGGAGCGCGGAGCGCG 1883  
QY 1797 GGTGCTCGACGAGACGCTTTCACCGCGGCGGCTTTCACCTTGAATATGCGCTCGC 1856  
DB 1884 CTGTGCTCGACGAGCGGCTTTCACCGCGGCGGCTTTCACCGGCGGCTTTCACCGGCTGCT 1943  
QY 1857 CGCGCTGTGGGCGGTGTGGGAGGTGTAGAGCGGAGGTGTGCGCGGCGGATGAGATCGGTGA 1916  
DB 1944 GCGGCTGTGGGAGAGCTGGGCGCTTGGCGCGCGGCGGAGGTGTGCGGCGGAGCTCGGCGCA 2003  
QY 1917 GGTGTGAGCTGCTGTGGGCGGCGGCGGCTTTCGCTTGAAGAGCGGCGGCTTTCGCTGG 1976  
DB 2004 ACTGACCGGCGCTTTCGCGGCGGCGGCGGCTTGTGTGCTGCGGCGGCTTTCGCGGCTG 2063  
QY 1977 TCGCGCGGCGGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2036  
DB 2064 CGCGCGGCGGCGGCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2123  
QY 2037 GCGGAGCGGAGGTGTGGGCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2096  
DB 2124 GACCGAGGAGCGAGGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2183  
QY 2097 GGTCAACGCTTCGAGCAGGTGTGTATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2156  
DB 2184 CGTCAACGAGCGGAGCGGT 2243  
QY 2157 GCGCGGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2216  
DB 2244 GCGCGGAGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2303  
QY 2217 CTACCGCTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2276  
DB 2304 CTTCGCGGAGCTGTAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2363  
QY 2277 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2336  
DB 2364 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2423  
QY 2337 GAGCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2396  
DB 2424 CCGGAGCGGCGGAGACATGTGGGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2483  
QY 2397 GAGGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2456  
DB 2484 CCGCGGCTTTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2543  
QY 2457 GCTGTGCTGT 2513  
DB 2544 GACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2603

QY 2514 GCGGCGTGGGCGTGAAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2573  
DB 2604 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2663  
QY 2574 CCGTGGCGT 2633  
DB 2664 CCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2723  
QY 2634 CAGGTAACCTTGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2680  
DB 2724 CAGGTAACCTTGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2770

RESULT 10  
US-09-266-965-74  
; Sequence 74, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456US1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 74  
; LENGTH: 12249  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
; US-09-266-965-74

Query Match 19.1%; Score 814.2; DB 4; Length 12249;  
Best Local Similarity 57.6%; Pred. No. 3.5e-134;  
Matches 1549; Conservative 0; Mismatches 1100; Indels 30; Gaps 4;

QY 18 CGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 77  
DB 1077 CGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1136  
QY 78 CGGTATGATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 137  
DB 1137 CGGTATGATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1196  
QY 138 AGTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 194  
DB 1197 GTTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1256  
QY 195 GGGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 254  
DB 1257 GGGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1316  
QY 255 CTTCCTTGGCATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 314  
DB 1317 CTTCCTTGGCATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1376  
QY 315 GAGGATGTGTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 374  
DB 1377 GAGGATGTGTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1436  
QY 375 GGAAGAGGAGGTGTTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 434  
DB 1437 GAGGATGTGTGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1496



QY 435 GACGCGCTCCGAGAGATGACGCTCANGCGGGCTGGGAGCAGATCCGAGCGTCCGAGC 494  
DB 1497 GGGGAGCGCTCCGCGAGGGGCTTC-----ATCGGAGCCGGCACTCCGACGATCCTCTC 1550  
QY 485 GGGCGGAACTTCGTATGCTCCCTGGGCTGGAGGGCCGTGTCTGGGGTGAATACGGCTTA 554  
DB 1551 GGGCGCGCTGCTCAGCTTGGGCTTCAGAGGGCCCGGCGGTGTCCGTGACACCGGCTG 1610  
QY 555 TTGCTCTCGCTGCTGAGCGGTTCACTGGGCTGTGAGAGCTTGCGCTCCGGGAAATGCTC 614  
DB 1611 CTCCTCTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1670  
QY 615 CACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674  
DB 1671 CTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1730  
QY 675 GAAAGACCCGCGGCGCTGGGAGGAGAGAGTGTGCTGCAAGGATTTTGGCGGAGCCGAGTGG 734  
DB 1731 CCGCGAGCGGGGCTGGGCGCGCGAGCGCGCTGCAAGTCTTCTGCGGACCGCGAGCGG 1790  
QY 735 GTTCCGAGCAGAGGCGAAGGGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 794  
DB 1791 CACCACTGTGTCCGAGGGGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1850  
QY 795 GAGCGGCGATGCGATATTGGCGGTGATTCGAGAGATCCGCGATCAATCAAGAGGTGCGAG 854  
DB 1851 CTTGGGCTAACCCCGTGCAGCGCGGTATCCGGGGCAGCGCGCTCAACAGAGAGGGCGAG 1910  
QY 855 CAGCGGCTGACACCGTGCAGAGGAGAGCTCCCAAGAAATGCTGTGAAGAGGGCCCTGGC 914  
DB 1911 CGCGGCGCTGACCGGCGCGCGCAAGAGACCGGCGCAAGCGGGGTGATCCGCGAGGCACTGCG 1970  
QY 915 GAGCGCAGGCTGCGCGCGCTTTCGGTGGGTTATGTGAGGAGCAAGGAGCGGGAGCAGC 974  
DB 1971 CAAGCAGCGCTGACGCGCGAGAGAGCTGACCGCGGTGAGGAGCAAGGAGCGGAGACCCG 2030  
QY 975 GCTTGTGATACCCCATGAAATCAAGCTCTGAATGCGGTATACGCGCTTCGCGAGAGATG 1034  
DB 2031 GCTGGCGACCGCGATGAGGCGCCAGGCGCTCTCGGACCTTACGGGGCGGGCGCGGAG 2090  
QY 1035 CGCCACGCGCTGCTGATGCGGTGCGGTGAAGACCAACTTGGCCATCTGAGTATGCGTC 1094  
DB 2091 GGGCGAGCGCTGTGCGGTGCGGTCTGCTGAAGTGAACCTGGGCGCACCACTGCGCGCG 2150  
QY 1095 GGGGATCACTGGGCTGCTGAAGGTGCTGTTCCTTCAAGCAGGAGAGATTCCTGCGCA 1154  
DB 2151 CGGCGCGCGCGCGCTGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2210  
QY 1155 CTTCCACGCGCAGCGCTGAGACCCCGGATCTCATGCGGTGATCTTTCGCTGACCGTAC 1214  
DB 2211 GCTGCACTCAAGGAGCCGACCGCGCGGTGATGATGATGATGATGATGATGATGATGATGAT 2270  
QY 1215 GCGCGCGCGGACACCGTGGCCGAGCTGGAATACGCGCGAGCGGGGGGTGAGCTGCT 1274  
DB 2271 GACCCAGGCGCCAGACTGGCGCGAGACCGGAGCGCGCGCGGTGCGGTCTGCTCT 2330  
QY 1275 CGGCAATGAGCGGGAGCAACGCGCAGGTGCTGGAAGAGGCGCGCGCGAGCACTGAGC 1334  
DB 2331 CGGCGCTCAGCGGCAACAGCCATGATGATCTTGAAGGCGCGCGCGCGAGAGAGCACC 2390  
QY 1335 ACCGCG-----GCGCGGAGCGAGCGGAGAGCTGCTGCTGCTGCT 1376  
DB 2391 GAGCGCGCGGCTGCGGAGCGTCTCTCGAGCGCGGGGGCGCGGTGCGGTGCTGCTC 2450  
QY 1377 GGCAGAGACCGGCTCAGCGCTTGAATGACAGGCGCGCGCTGCGGACCATCTGAGAGC 1436  
DB 2451 CGGCGCGAGCGAGGCGCGCTGCGAGCGAGGCGAGCGCTGCGCGCGCGCACTTACCGC 2510  
QY 1437 CTACGCTTCCGAGTGTGCGGAGATGAGCTTTCAGTCTGCGAGAGAGCGAGGCGGAT 1496  
DB 2511 GCGCGCGAGCTGAGCACCGGCGAGGAGTGCACCGCGCTGAGCAACGCGGGCGGCTT 2570  
QY 1497 GAGAGCAGCGGCTGCGGTGCGGCGAGCGTGAAGGAGGGGCTGCGGCGAGCCCTTGAAGC 1556

DB 2571 CGACCAAGGCGCGCGCTGCTGCGGCGGAGCCGTGAGAGAGCTGCGCGCGCTGCGCGG 2630  
QY 1557 TCGCGCGAGGAGACAGAGTGGCGCGCGCGGTGGCGAGTATCCGCAATTCCTCACGGGG 1616  
DB 2631 CTTGCGAGCGGAAACCGCGCGCGCGCTGTATCCGCGGAGCCCGCGCTCGCGGG 2690  
QY 1617 CAAGCTCGCTTCTTCTTCAACCGGACAGGAGGAGCGAGAGCTGAGGATGAGGCGTGGGCT 1676  
DB 2691 CAAGCGCGCTTCTTCTTCAACCGGACAGGAGGAGCGAGAGCGCGGAGATGAGGCGGAACT 2750  
QY 1677 GTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1736  
DB 2751 GAGCGCTCAGACAGACCGGTGCTGCGGAGCGCTGAGAGAGGATGATGATGATGATGATGAT 2810  
QY 1737 GAGCTGCAACCGCGCGCTCCGCGAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 1796  
DB 2811 GACTCTGACCGCGCGCTGCGCGAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 2870  
QY 1797 GCTGCTCAGCAGACAGGCTTCAACCGAGCGCGGCTGTTCACTTGAATATGCGCTGCG 1856  
DB 2871 CTTGCTGACAGAGAGGCTTCAACCGAGCGCGGCTGTTCACTTGAATATGCGCTGCGCT 2930  
QY 1857 GCGCTGTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1916  
DB 2931 GCGGCTGTGAGAGAGCTGAGGCTTGGCGCGCGGATGCTGCGGCGGCACTGCGTGGCG 2990  
QY 1917 GCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1976  
DB 2991 ACTGACTGCGCTTCAACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3050  
QY 1977 TCGCGCGGCGCGCTGATGAGGCGCTGCGCGCGCGCGGAGATGATGATGATGATGATGAT 2036  
DB 3051 CGCGCGCGCGCGCTGAGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3110  
QY 2037 GCGGAGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2096  
DB 3111 GACCGAGAGCAGAGGTGCG 3170  
QY 2097 GGTCAAGCTTCCGAGCAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2156  
DB 3171 GGTCAAGCAGAGCAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3230  
QY 2157 GCGGCGAGTGGCG 2216  
DB 3231 GCGGAGTGGCG 3290  
QY 2217 CTACCGCTCATGCG 2276  
DB 3291 CTTGCGCGACCGTGAAGCGCATGACCGAGGCGCTTGGCGAGGTGCGAGAGGAGGTGCTGTA 3350  
QY 2277 CCGGCGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2336  
DB 3351 CAGGCGCGCGCTTCTTCCGAGGTCTTCAACCTCAACCGGCGCGCGCGCGCGCGCGCGCG 3410  
QY 2337 GAGCTGCGCGGAGTATGAGGTGCGCGCAGCGCGAGAGGTGAGGCTTCCGCGAGTGAAGT 2396  
DB 3411 CCGGAGCGGAGAACTGAGGTGCGGACGCTCCGAGAGAGGTGCGCTTCCAGAGCGGCT 3470  
QY 2397 GAAAGCGCTGACGCGCGCGGTGCGGAGCACTTGTGAGGTGAGTCCGAATGACGCT 2456  
DB 3471 CCGGCGCTTGGCGAGCGCGGAGCGACCGCGCTTCTGAGAGGTGCGGCGCGCGCGGCTGCT 3530  
QY 2457 GCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2513  
DB 3531 GAGGCGCGCGCAGCGCGAGTCTTGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3590  
QY 2514 GCGCGCTGAGGTGAGAGCT 2573  
DB 3591 GCGGCGCGGAGCGCGGAGCGCGAGTCCGTCTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3650  
QY 2574 CCGTGGCTGTCTCTGCGCGCGCGCTTCTTCCCTCAGAGGAGCGCGCGGCTGCGCGCTGCG 2633

Db 3651 GGGCTCTGCGCGGGTGGGACCGCTGCTGCCCAAGCCCGGAGCGCGGTGACCTGCC 3710  
Qy 2634 CACGTACCTTGGCAGCGCGGAGCGCTACTGTGATCGACACGAAAGCCG 2680  
Db 3711 CACGTACCGCTTCCAGCGCGGCACTACTGCTGGCGGAGCGGACG 3757

RESULT 11  
US-09-266-965-96  
; Sequence 96, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600,456US1  
; CURRENT APPLICATION NUMBER: US/09/266,965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 96  
; LENGTH: 18311  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
; US-09-266-965-96

Query Match 19.1%; Score 814.2; DB 4; Length 18311;  
Best Local Similarity 57.6%; Pred. No. 3,6e-134;  
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

Qy 18 CGAGCGCGGAGCGGAGATCCGATTCGATCGTCGAGCGGAGTGGCCGTCGCCGGTGG 77  
Db 1077 CGAGACGCGGCGCGCGAGCCCGTCGCGATCGTGGGCGATGGCGGCTTCCCGGGGA 1136  
Qy 78 CGTGATCATCTGACGCGGCTTCTGACGCTCTCCAGGGCTGGCGGAGACCGTGGGG 137  
Db 1137 CGTGACATCGCGAGACCTGTGTGCGTGGCGAGGCGCGGAGCGCGTACCGA 1196  
Qy 138 AGTCCCGCGGAAAC--GCTGGAGTGCAGACGCTGTGTTGATCCGACCCCGATGCC 194  
Db 1197 GTTCCCGCGGACCGGGGCTGGAGCGTGCACGCCGCTACGACCCCGAGCCGACCC 1256  
Qy 195 GGGGAAGACGCCGCTTACCGCGCATCTTCTGAGCGAGTACCTGCTTGCAGCGCTC 254  
Db 1257 GGGGAGAGCTAGCGCGCGCACGCGCGCTTCTCAAGACGCGCGGATTCGAGCGCG 1316  
Qy 255 CTCTTTCGCACTCGCTCGCGAGCGCTGCGAGTGCAGTGCATCGACTCTTGT 314  
Db 1317 CTCTTCGCACTACGCGCGCGAGCGCTGCGCATGCAGCGACGACGATGATCAT 1376  
Qy 315 GAGAGTGTGCTGGAGAGCGCTGAGAACCGCGCATCGCTCCATCGCGCTGTGCTAC 374  
Db 1377 GAGAGTGTGCTGGAGAGCGCTTGAAGAGCGCGCTTCAAGACGACCACTCGGGGCGA 1436  
Qy 375 GGAAGAGCGAGTGTTCATCGGAGTGGCGCTTCGAAATGAGAGCGCGCTGCGCAAC 434  
Db 1437 GAGAGTGTGCTGTGCTGTGCTGCTCAACAGCAAGCATCTGATCAAGCTGTGAGCG 1496  
Qy 435 GACGCGTTCGCGAGATGAGACGCTCATGAGCGGCGCTGGGAGAGATCCAGAGTGGAG 494  
Db 1497 GCGGAGCGTGGCGAGAGGCTTC-----ATCGGAGCGGCAACTCGCGAGATCTCTTC 1550  
Qy 495 GGGCGGAATCTGTATGCGCTCGGCGTGGAGGCGCGTGTGTGCGAGTATACGAGCTTA 554

Db 1551 GGGCGCGTTCGCTACACCTTTCGCTTCAGAGGCGCGCGCTGTCTGACACCGCGCTG 1610  
Qy 555 TTGCTCTGCTGTGTGCGCGCTTTCATCTGCTGTGTGAGCTTGGCTCCGGGAAATGCTC 614  
Db 1611 CTCTCTCTGCTGTGTGCGCTGTGCTGCTGCTGCGCGAGTCCCTGCGGAGGAGTGTCTC 1670  
Qy 615 CACGCGCTGCTGTGTGAGTATGCTGATGTGTCGCGAGACCTCGTGTGCTCTC 674  
Db 1671 CTGCGCTGTGCGCGCGCGCGAGCGATGAGCCAGCGACCGCTTTCATGATGATCAG 1730  
Qy 675 GAAAGCCCGGCGCTGCGCAAGAGAGTGTGCTGCAAGGCACTTTCGCGAGGCGGATG 734  
Db 1731 CCGCAGCGGCGCTGCGCGCGCGCGAGCGCGCTGCAAGTCTTCTGCGGAGCGCGAGCG 1790  
Qy 735 GTTTCGAGAGAGCGAGGCTGCGCTGTGTGCTCTCAAGGCGCTCACTGAGAGCCCGCG 794  
Db 1791 CACCACTGTGTCCGAGGCGCGCGCTGTGCTGCGCGCTGCGAGCGCGCGCG 1850  
Qy 795 GAGCGCGATCGGATATTTGGGCGGTGATTGAGGATCCGCGATCAATCAAGCGGTCGAG 854  
Db 1851 CTGCGCTACCGCTGTGCAAGCGGTGATTCGGGCGAGCGCGCTCAACAGAGCGCGAG 1910  
Qy 855 CAGCGCTGACCGTGTGCGAAGCGAGAGCTCCCAAGAAATGCTGTAACCGGCGCTGAG 914  
Db 1911 CCGCGCGCTGACCGCGCGCGCAAGAGCGCGCGCGCAAGCGGAGTATCCGCGAGGCACTG 1970  
Qy 915 GAGCGAGGCTGCGCGCGCTCTTCTGCTGTGTATGTGAGGCAACCGCGAGGAGAGAC 974  
Db 1971 CAACCGACGCGTGAACGCGCGAGAGAGCTGAGCGCGCTCAGGAGCAACCGCGACCG 2030  
Qy 975 GCTTGTGACCGCATCGAAATCGAAATCGTGAATGCGGATACGCGCTCGGCGAGATGT 1034  
Db 2031 GTTGTGCAACCGATCGAGCGCGCGAGCGCTCTCTGCGACCTACAGCGCGCGCGCGA 2090  
Qy 1035 CGCGACCGCGCTGTGATCGGCTGTGAGAGCAACACCTTGGCCATCTGATGCTGCTC 1094  
Db 2091 GGGGAGCGCGCTGTGCTGTGCTGTGCTGTGAGTGAACCTGAGGCGCAACCGATCGCG 2150  
Qy 1095 GGGGATCACTGAGCTGTGAGAGTGTGTCTTCTTCAAGCAACGCGGAGATTTCTGCGCA 1154  
Db 2151 CGGCGCGGCGCGCTGATCAAGATGTATGAGCAATGAGGAGCGAGCGCTGCGCGAGC 2210  
Qy 1155 CTTCCAGCGAGGCGGTGAAACCGCGGATCTCATGTGGTATCTTGGCTGACCGTCA 1214  
Db 2211 GTTCACTTCAAGAGCCGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2270  
Qy 1215 GCGCGCGGAGCAACCGTGGCGGAGTGAATACCGCGGAGCGGCGGAGTGAAGCTGCT 1274  
Db 2271 GACCGAGGCGGAGACTGGCGGAGCAACCGGAGCGCGCGCTGCGCTGCTGCTGCT 2330  
Qy 1275 CGGCAATGAGGAGCAACCGGCAAGTGTGTGTGTAAGAGCGCGCGGCGGAGCTGCA 1334  
Db 2331 CGGCGTCAAGCGGCAACCGCGCATGTATCTTGAAGGCGCGCGCGCGGAGAGGCAAC 2390  
Qy 1335 ACCGCG-----GCGCGGAGGAGCGGCAAGAGCTGTGCTGTCTC 1376  
Db 2391 GAGCGCGCGCTGCGGAGAGCTCTCTGCGAGCGCGGAGCGCGCTGCTGCTGCTGCT 2450  
Qy 1377 GGCAGAGACCGCGTACCGCTGATGACAGCGCGCGCGGCTGCGGACCATCTGAGAC 1436  
Db 2451 CGGCGGAGCGAGGCGCGCTGCGGAGCGGAGCGGAGCGGCTGCGGAGCGGAGCGG 2510  
Qy 1437 CTACCTTTCGAGTGTCTGCGGAGTGTGCGCTTCAAGTCTGCGAGAGCGGAGCGGAT 1496  
Db 2511 GCGCGCGGACCTGAGCAACCGGCGAGTGTGCGGACCGCGCTGCGGAGCGGAGCGGCT 2570  
Qy 1497 GAGAGCGGCTGCGGAGTGTGCGGAGCGTGCAGGAGGAGGCTGCGGAGCGGCTGAGAG 1556  
Db 2571 GAGAGCGGCGCGCGCTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2630  
Qy 1557 TCGCGCGAGGAGCAAGAGTGTGCGCGGAGTGTGCGGAGTGTGCGGATTCCTCAAGCG 1616  
Db 2631 CTGCGCAACCGGAGCGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2690

1617 CAAGCTCCCTTTCTTCTTACCGGACAGGGGCGAGACGCTGGGCATGAGCCGTGGGCT 1676  
Db CAAGGCGCGCTTCTTCTTACCGGACAGGGGCGAGACGCTGGGCATGAGCCGTGGGCT 2750  
Qy 1677 GTACGATGTATGATGCTCCGCTTCCGAGGCGCTTCACTTGTGCTGAGGCTGTTCACCA 1736  
Db 2751 GCGCGCTACAGACCGGTGTTCGCGAGCGCTTGAACAGGTTCGCGCCAGCTCGACCG 2810  
Qy 1737 GGAGCTGACCGGCGCTCCGAGAGGTATGAGGCGCAACCGGCGAGGTCGACCGCG 1796  
Db 2811 GCACCTGACCGGCGCTCCGAGAGGTGTTCGCGCGAGGCGAGCGCGCGCGCG 2870  
Qy 1797 GCTGCTGACGAGACGCTTTCACCGAGCGCGCGCTGTTCACCTTGAATATGCGCTCG 1856  
Db 2871 CTTGCTGACAGACGCGCTTACACCGAGCGCGCTGTTCGCGCGAGGTCGCGCTCGT 2930  
Qy 1857 GCGCTGTGAGGCTGTGAGGCTGAGACCGGAGTGTGCTGCGCGCGCATGACATGCTGA 1916  
Db 2931 GCGCTGTGAGGACTGAGGCTGTGCGCGCGCATGAGTGTGCGCGCGCATGCTGCGCG 2990  
Qy 1917 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1976  
Db 2991 ACTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3050  
Qy 1977 TGCGCGCGGCGCGCTGATGACGCGCTGCGCGCGCGCGCGCGCGCTGCTGCTGCTG 2036  
Db 3051 CGCGCGCGGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTG 3110  
Qy 2037 GCGCGAGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2096  
Db 3111 GACCGAGGACGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3170  
Qy 2097 GGTAAAGCTCCGACCGAGTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2156  
Db 3171 GGTAAAGCTCCGACCGAGTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3230  
Qy 2157 GCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2216  
Db 3231 GCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3290  
Qy 2217 CTACCGCTCAATGCG 2276  
Db 3291 CTCCG 3350  
Qy 2277 CCG 2336  
Db 3351 CAGCG 3410  
Qy 2337 GAGCTCG 2396  
Db 3411 CCGCGAGCGCGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3470  
Qy 2397 GAAAGCGCTGACCG 2456  
Db 3471 CCG 3530  
Qy 2457 GCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2513  
Db 3531 GACG 3590  
Qy 2514 GCGCGCTGCG 2573  
Db 3591 GCG 3650  
Qy 2574 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2633  
Db 3651 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3710  
Qy 2634 CACGTAACCTTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2680  
Db 3711 CACGTAACCTTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3757

RESULT 12  
US-09-144-085-3  
; Sequence No. 6280999  
; Patent No. 6280999  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Claes  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Ashley, Gary  
; APPLICANT: Julien, Bryan  
; APPLICANT: Ziemann, Rainer  
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
; FILE REFERENCE: 30062-20020.20  
; CURRENT APPLICATION NUMBER: US/09/144.085  
; EARLIER FILING DATE: 1998-08-31  
; EARLIER APPLICATION NUMBER: 09/010.809  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 33529  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-09-144-085-3  
Query Match 18.8%; Score 803.2; DB 3; Length 33529;  
Best Local Similarity 57.6%; Pred. No. 3.1e-112;  
Matches 1591; Conservative 0; Mismatches 1128; Indels 41; Gaps 7;  
Qy 37 CCGATTGCGATGTCGAGGAGCGAGTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 96  
Db 25316 CCGATTGCGATGTCGAGGAGCGAGTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 25375  
Qy 97 TTCGAGCGCTCTCGAGAGCGCTCCGCGAGACACCGTGGCGAGTTCCTCCGCGAGCGCTG 156  
Db 25376 TTCGAGCGCTCTCGAGAGCGCTCCGCGAGATACGCTGCGCGCGCGCGCGCGCGCG 25435  
Qy 157 GATGAGGAGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 216  
Db 25436 GATGAGGAGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25495  
Qy 217 GCATCTTTCCTGAGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 276  
Db 25496 GCATCTTTCCTGAGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25555  
Qy 277 GAAAGCGCTGCGAGGAGCGCTGACATGACTCTTCTGAGAGGTGTGCTGGAGCGCTG 336  
Db 25556 GAAAGCGCTGCGAGGAGCGCTGACATGACTCTTCTGAGAGGTGTGCTGGAGCGCTG 25615  
Qy 337 GAGAACGCGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396  
Db 25616 GAGAACGCGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25675  
Qy 397 ATCGCGCGCGCGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 456  
Db 25676 ATCGCGCGCGCGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25726  
Qy 457 GCTATGCGCGCGCGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516  
Db 25727 GCTATGCGCGCGCGCGAGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25786  
Qy 517 GGGCGCGAGGCGCGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 576  
Db 25787 GGGCGCGAGGCGCGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25846  
Qy 577 CATCTGCGCTGTCGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 636  
Db 25847 CATCTGCGCTGTCGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25906  
Qy 637 TCGCTGATGTTGTGCGCGAGACCTGCTGTGCTGCTGCGAGACCGCGCGCGCGCGCG 696  
Db 25907 TCGCTGATGTTGTGCGCGAGACCTGCTGTGCTGCTGCGAGACCGCGCGCGCGCGCG 25966

QY 697 GACGCTGCTGCAAGGCAATTTTCGCGGAGGCGGATGAGTTCGGAGACGAGCCGAAGGATGC 756  
Db 25967 GACGGGCGGGTGCAAGCGCTTCTCGACGGGCGGAGGATTATGCGCGTGGCCAGAGGATGC 26026  
QY 757 GCGCTGCTGATCTCAAGCGGCTCAGTGTAGGACCGCGCGGAGCGGATGAGTATTTAGGCG 816  
Db 26027 GGGGTGCTGATGCTGATGCGGCTGTCGAGGCGGAGCGGAGGAGGCGGCTGCTGCTGCT 26086  
QY 817 GTGATTTCAGATTCGCGATCAATCAACGAGGTCGAGCAGCGGCTGACCGTTCGAGC 876  
Db 26087 GTGTGCGGCGGACGCGCGGCTCATACGAGCGGCGGAGCAGCGGAGTCAACGCGCGCGAAC 26146  
QY 877 GGGAGCTCCCAAGAAATGCTGCTGAACCGGCGCTTGCGGAGCGGAGGCTGCGCGCTGT 936  
Db 26147 GGCACGCGCGCACAGAAAGGTGTGTGCGCGGCTGCGGAGCGGAGGCTGAGGCGCGCG 26206  
QY 937 TCGGTGGATTATGTGAGGACACAGGCGGAGCGGAGCGGATGTCGCAACGCGCTGCTGATCGG 996  
Db 26207 AGCATGATGTGTGTGAGTGCACGAGTACGAGCATGCTGCTGAGGCGACCGCATCGAGGTG 26286  
QY 997 CAAGCTCTGAATGCGGTATACGAGCTCGGCGGAGATGTCGCAACGCGCTGCTGATCGG 1056  
Db 26267 CAGGCGCTCGGCGGCTGTACGAGGCAAGGTGCGGATATGCTGCTGCTGAGCTGAGGCTG 26326  
QY 1057 TGGGTGAGAACCACTTGGCCATCTGAGTATGCTGCGGAGATCACTGGGCTGCTGAG 1116  
Db 26327 GCGGTCAAGAAATATCGGTATCTGAGTCCGCGGAGATCGCAAGGAGGTGTGCAAG 26386  
QY 1117 GTGCTTGTCCCTTCAGACGAGGCGAGATTCCTGCGCACCTTCACGCGGAGCGCTGAGAC 1176  
Db 26387 ATCTGTGCGGCGCTTCCGTTACGAGTGTGCTGCGGAGACGCTGACGCTGCGCGCAAT 26446  
QY 1177 CCCCAGATTCATGAGGTGATTTTGGCTGTACCGCTGACGCGCGCGGACACCGTGGCC- 1235  
Db 26447 CCCCCATCCCTGGGAGAACTGCGCGGTGCAAGTGTGATGCTGCTGACCCCTGAGCCT 26506  
QY 1236 -----GACGTGAATACCGCGGAGCGGCGGAGGTGAGTGTGCTTGGAGTGAAGCGGAGC 1290  
Db 26507 CCGCGCGAGAGAGGCGCGCGCGGTGCGCGGTGTGCTGCTGCGGATACGCGGAGC 26566  
QY 1291 AACGCGCATGCTGTGAGAGAGGCGCGGCGGAGCGGATGCAACCGCC-----CG 1342  
Db 26567 AACGCGCATGCTGTGAGAGAGGCGCGCGGAGCGGCGGAGCGGATGCAACCGG 26626  
QY 1343 CCGCGGAGGACCGGCGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1402  
Db 26627 AGGCGGCGCTGCTGCTATTTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 26686  
QY 1403 CACACGCGCGGCGGCTGCGCGGACCATCTGTGAGACCTACCTTCGACGTGTGTGTGTGTGT 1462  
Db 26687 CCGAGGCGGCGGCGGT 26746  
QY 1463 TGGCTTTCAGTCTGCGGACGAGCGCGGATGAGACCGGCTCGGCTGTGTGTGTGTGTGTGTGT 1522  
Db 26747 TGGTGTGCGACGCGGCGGCTGT 26806  
QY 1523 CGTTCAGAGGAGGCGGCTGT 1582  
Db 26807 CAGAGGCTGT 26866  
QY 1583 GTGCGGT 1642  
Db 26867 CCGGT 26923  
QY 1643 AGGCGGCGGACGCTGT 1702  
Db 26924 AGGCGGCGGACGCTGT 26983  
QY 1703 AGGCGTTCAGCTGT 1762  
Db 26984 CCGGCTTCAGCTGT 27043

QY 1763 TGT 1822  
Db 27044 TGT 27103  
QY 1823 AGCGGCGGCTGT 1882  
Db 27104 AGCGGCGGCTGT 27163  
QY 1883 AGCGGAGT 1942  
Db 27164 AGCGGAGT 27223  
QY 1943 TGT 2002  
Db 27224 TGT 27283  
QY 2003 TGT 2062  
Db 27284 TGT 27343  
QY 2063 TGT 2116  
Db 27344 TGT 27403  
QY 2117 TGT 2176  
Db 27404 TGT 27463  
QY 2177 TGT 2236  
Db 27464 TGT 27523  
QY 2237 TGT 2296  
Db 27524 TGT 27583  
QY 2297 TGT 2353  
Db 27584 TGT 27643  
QY 2354 TGT 2413  
Db 27644 TGT 27703  
QY 2414 CCGGT 2473  
Db 27704 CCGGT 27763  
QY 2474 CCGGT 2527  
Db 27764 CCGGT 27823  
QY 2528 TGT 2587  
Db 27824 TGT 27883  
QY 2588 TGT 2647  
Db 27884 TGT 27943  
QY 2648 TGT 2707  
Db 27944 TGT 28003  
QY 2708 TGT 2767  
Db 28004 TGT 28063

RESULT 13  
US-08-804-227C-1  
; Sequence 1, Application US/08804227C  
; Patent No. 5876991

GENERAL INFORMATION:  
APPLICANT: DeHoff, Bradley S.  
APPLICANT: Kuhse, Stuart A.  
APPLICANT: Rostack, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43280 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 816..14234  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14351..19945  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20010..31199  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31232..36067  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36249..41774  
US-08-804-227C-1

Query Match 17.7%; Score 755.2; DB 2; Length 43280;  
Best Local Similarity 56.0%; Pred. No. 7.8e-124;  
Matches 1603; Conservative 0; Mismatches 1208; Indels 53; Gaps 7;

QY 24 CGCGCCGAAGATCGATTGCGATGCTGCGAGCGAGTGGCCGCTGCCCCGCTGGCGTAT 83  
DB 36350 CGCGCCGCGGAGCGGATGCGCTGCTGCGAGCGAGTGGCCGCTGCCCCGCTGGCGTAC 36409  
QY 84 CGATCTGACGCGGTTCTGAGCGCTCTCGAGGCTCGCGCGACCGATCGGCGAGTCCC 143  
DB 36410 CGGCGCCGAGAGAGCTGTGGAGCTGTGGCCGAGAGCGGAGCGGATCGGCGCTTCCC 36469  
QY 144 CGCGGAGC--GCTGGAGTGAAGAGCGGTGTTGATCCGACCCCGATGCCCCGGGAA 200  
DB 36470 CGTGGACCGGGGCTGTGGAGCTGTGGCGTGGATCGACCCGATCCGAGTGGAGGAC 36529  
QY 201 GACGCGCTTACGCGCGCATCTTCTGAGCGAGCGTACCTGCTTCCGAGCTCTTCTT 260  
DB 36530 CACGTAAGCGCGGAGCGGCTTCTGAGAGCGCGCGTGAATTGACGCGCTTCTT 36589

QY 261 CGGCATCTCGCTTCGCGAAGCGCTGCGAGTGAACCTTGACATGACTCTTGTGAGAGT 320  
DB 36590 CGGCATCTCGCTTCGCGAAGCGCTGCGAGTGAACCTTGACATGACTCTTGTGAGAGT 36649  
QY 321 GTGCTGAGAGCGGTTGAGAGAGCGCGCGATGCTTCAATCGCGCTGCTGCTGAGGAAAC 380  
DB 36650 GTCTGAGAGCGGTTGAGAGAGCGCGCGATGCTTCAATCGCGCTGCTGCTGAGGAAAC 36709  
QY 381 GGAAGTTCATCGGAGTGGAGCCCGTCCGAATATGAGGCGCGCTGCGCAAGCGAGCGC 440  
DB 36710 TGATGTCTACGTGGAGCGCGCGAGCGAGCTGTAGCGCTCGATCCCGCTGTGTCCGA 36769  
QY 441 GTCCGAGAGATGACCGCTCAATGCGCGGCTGAGGAGCAATGCGAGCTGAGAGCGGCGG 500  
DB 36770 GGGCTCGGAGGCTATGCTGAGCGGAGCGGAGCGGCGG-----GTGATGTCCGCGG 36823  
QY 501 AATCTCGTATGCTTGGGCTGCGAGGCGCTGTGTGCGGTTGAGTACGGCTTATGCTC 560  
DB 36824 CATCTCTTACGCGCTGCTGCTGAGAGAGCGCTCAATGAGCGGTGAGACGCGCTGCTCTC 36883  
QY 561 CTGCTGAGTGGCGGTTGATGCTGATGCTGCTGAGAGCTGCGCTGCGGAGTCTCCACGCGC 620  
DB 36884 CTGCTGAGTGGCGGTTGATGCTGATGCTGCTGAGAGCTGCGCTGCGGAGTCTCCACGCGC 36943  
QY 621 CTGCTGAGTGGCGGTTGATGCTGATGCTGCTGAGAGCTGCGCTGCGGAGTCTCCACGCGC 680  
DB 36944 GCTGCGCGGCGGCGGTTGAGAGAGCGGATGCGGAGTCCGCGGCGTTGAGAGTCTCCGCGCA 37003  
QY 681 CGGCGCGTGGCGAGAGAGCGTCTGCTGAGAGCATTTTCCGCGAGGCGGATGAGTTGG 740  
DB 37004 GAAAGGCGTGGCGCGCGAGCGGCTGCTGAGAGCGGCTTCCGCGCGCGCGCGAGCGAGC 37063  
QY 741 ACGAGCGAAGAGGAGCGCGCGCTGCTGAGAGCGGCTGAGAGCGGCGCGAGCGAGC 800  
DB 37064 CTGCGCGGAGGCGGCTGCGGCTGCTGCTGAGAGCGGCTGAGAGCGGCGCGAGCGAGC 37123  
QY 801 CGATCGATATGCGCGGTGATTCAGAGATCCGATCAATGAGCGGTGAGAGCGAGC 860  
DB 37124 GCAACAGGTCTCTGCGCTGCTGAGAGCGGCGGCTGAGAGCGGCGGCTGAGAGCGG 37183  
QY 861 TCTGACCTGCTGAGAGAGAGAGCTCCCAAGAAATGCTGAGAAAGGCGCTGAGAGAGC 920  
DB 37184 GCTGACCGCGCGCGAGAGAGAGAGCTCCCAAGAAATGCTGAGAAAGGCGCTGAGAGAGC 37243  
QY 921 AGGCTGCGCGCGCTTCTGAGGTTATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGC 980  
DB 37244 CGGCTGCTCCCGAGC 37303  
QY 981 TGACCCGATCGAATTCAGAGCTTGAATGCGGTATGAGCGCTCGGCGAGATGTCGACAC 1040  
DB 37304 CGACCCGATCGAATTCAGAGCTTGAATGCGGTATGAGCGCTCGGCGAGATGTCGACAC 37363  
QY 1041 GCGCTGCTGATCGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1100  
DB 37364 CCCCTGCTGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 37423  
QY 1101 CACTGAGGCTGAGC 1160  
DB 37424 CGGCGGCTGATCAAGAGTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 37483  
QY 1161 CGCGAGAGCGTGAACCCCGGATCTCATGAGGAGTATCTTGTGAGTACCGTCAAGCGGCGC 1220  
DB 37484 CGCGAGAGCGTGAACCCCGGATCTCATGAGGAGTATCTTGTGAGTACCGTCAAGCGGCGC 37543  
QY 1221 CGGAGACAGTGGCGCGAGC 1280  
DB 37544 CGAGAGCGTGGAGC 37603  
QY 1281 GAGGAGC 1340  
DB 37604 CGGCGAGC 37663  
QY 1341 GCGCGCGAGC 1367

[illegible]

Db	38741	CGCGGACGGCGGTCCGACCGGCACACGGCTCCGAGGCCCGCACCTTCTCTGAGACCGGCCC	38800
QY	2445	GAATTCGACCGCTCTGGGCTCTGGTGGCTCTCCATGACCGGAGCGCCCGGCGCTGTCT	2504
Db	38801	GGAGGCGGCTCTGTGTGGCGGATGGCAGAGAGATGCTCTGGAGAGCGACACCTGAGCCCTGTCT	38860
QY	2505	CGCATCTGTCCGCGCTTGGCGGCTGATCAGACCGGCGGACCTGTGCTTCAGAGCGGCTCTGGCGGCT	2564
Db	38861	---GCCCGGATCCACAAAGCCGGGACCGGGCCGACCGGTCCGGGGGCTCCGGGGCGGT	38917
QY	2565	CTGGGCGCTGTGGTGGCTGTGTTCTCTGGGCGCGGCTCTTCCCTCAGAGGGGCGGCGGT	2624
Db	38918	GCGGGCGGCGCGCGCGCTGACCGGCGGGGCGCCCGGATGACT--GGGCGCGGATGCAC	38975
QY	2625	GCGGCTGCCACGACACCTTTGGGAGCGCGAGCGGCTATCGATGCACAGAAAGCCGACGA	2684
Db	38976	GCCACGAGCCCCCGAGGGGCGGAGCCCGCGCGCTGTGAATCTCCGCTCAGCGCTTCCGGCAC	39035
QY	2685	CGCGGCGGTCGGCGGACCGCGGCTCTCCGGGAGCGGCTCAGCAGAGTTCGAGAGAGGGGG	2744
Db	39036	CGCGGCTACTGGCTCGGCGCCCGGGGCGGCGGCGGACACGAGCGACTGATTAACGGATTC	39095
QY	2745	CGCGGTCGCGGCGGCGGACCCGGCGGACGCTCGGCTTCAGCAATCCGCGCGCCGAGAGCGG	2804
Db	39096	GGCTGGGACCGGCTGCGCGCTGTGAACCGGCGGGGCCCGGACCGCGCGGCTGGCTGGTG	39155
QY	2805	ACGCGGGAGAGAGTGGAGGCGCGCGGCGGACCGTCCGTTCCGGC	2848
Db	39156	ATTCACCCGACACCGCGCGCTCCGGAGACTGTCTCGGCGACAGC	39199

RESULT 14  
US-09-105-537-1  
; Sequence 1, Application US/09105537A

```

; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, P.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

Query Match	17.4%;	Score 742;	DB 3;	Length 15872;
Best Local Similarity	-55.7%;	Pred. No. 1.5e-121;		
Matches 1594;	Conservative	0;	Mismatches 1215;	Indels 53; Gaps 7;

Qy	26	CAGCGGAAGATCCGATTTGCATTCGTCGAGGAGATGCGCTTCGCGCGGATGATCG	85
Db	8490	CCGACGACGATCCGATCGCGATCTGTGGGACATGGCATGCCCTATCCGGGTGTGTGCT	8544
Qy	86	ATCTGAGCGGGTTCTGACGCTCTCTCGAGGCTCGCGACACCGTCGGCGAGTCCCG	145
Db	8550	CGCGGAGAGACTGTGGCGCGCTGGTGGCCGAGGGGACGAGACGCGATTCAGCGATGCCCG	8605
Qy	146	CCGAACT---GCTGGGATTCGACAGCGTGTGATTCCGACCCCGATGCCCGGGGAGA	202
Db	8610	TCAACTCGCGGCTGGGACTCTGAGAGCCTCTACGACCCGGAATCCGAGTTCGAAAGGACA	8665
Qy	203	CGCCCGTTACCGCGCATCTTTCCTGACGCGAGTAGCCTGTCTCGACGCGCTCTTCTTCG	262
Db	8670	CGTACTGCGGGAGGGGCGGGTTCTCGAAGAGGCGCGGTGACTTCGACGCGCGCTTCTTCG	8722
Qy	263	GCATTCGCGCTTCGCGAGCGCTCGCGGATGAGACCTTCGACATTCGACTTTCCTCGAGAGTCT	322

Db 8730 GCACTCGCCGCGAGGAGCCCTGTGATGGAACCCGACAGCCGAGCTGCTGTGAGAGTGT 8789  
Qy 323 GCTGGAGGCGCTGAGAAAGCCCGATCGCTCAATCGGCGCTGCTGTGATGAGAAACGG 382  
Db 8790 CTTGGAGGCGCTGAGAAAGCCCGATCGAATCGCTGCTGTGAGAGGCGCGCTG 8849  
Qy 383 GAGTGTTCATCGGAGTGGCCCGCTCGAATATGAGAGCGCGCTGCGGAGAGCGAGCT 442  
Db 8850 GTGTCTAGTGGGCGCGCGACCGCTGTGACGCTCCGATCCCGGCTGTGCTCCAGG 8909  
Qy 443 CGGAGAGATGACGCTCATGCGGCGCTGCGGAGAGTCCAGCGCTGCGAGCGGCGGA 502  
Db 8910 GCTCGGAG-----GGCTATCTGCTGACCGGACGCGGAGCGGAGTGTCTCCGCGCA 8963  
Qy 503 TCTGTATGCGCTGCGGCGTGGAGAGCGGCTGTGTCGAGTGAATACCGGCTATCTGCT 562  
Db 8964 TCTCTATGCGCTGCTGTGTCGAGAGCGCTGTCATGAGTGAAGAGCGGCTGTCTCT 9023  
Qy 563 CGCTGTGGCGCTTCATCTGAGCTGTGACAGCTTGCGCTCGGAGATGCTCAAGCGCC 622  
Db 9024 CGCTGTGGCGCTGTGATGTGGCGGATACGGGCGCTGCGGAGAGTGGCGGCTCGCG 9083  
Qy 623 TGGCTGTGGGATCGCTGATGTTGTGCGCGAGACCTCTGTGCTCTCGAAGACC 682  
Db 9084 TGGCGGCGGCGGCTGTGATGCGCGATCGCGCGCGCTGTGAGATCTCCGCGCAGA 9143  
Qy 683 GGGGCGTGGCGAGGAGCGCTGCTGACAGGATTTTCGCGGAGCGGATGAGTTTCGAC 742  
Db 9144 AGGGGCTGGCGCGGAGCGGCTGACAGGAGTTCGCGCGCGCTGACAGCGAGCT 9203  
Qy 743 GAGGCGAAGGCTGGCGCGCTGTGCTCAAGCGGCTGATGAGCGCGCGCGAGCGAG 802  
Db 9204 GGGCGAGGCGGCTGTGCTGTGCTGTGAGCGGCTGTGAGACCGCGCGCGCGGCGG 9263  
Qy 803 ATCGGATATGCGGATGATGAGAGATCGCGATCATACGAGCGGTGCGAGCGGCTC 862  
Db 9264 ACAGGCTCTCGGCTGTGACCGGACCGCGGCTGACAGAGAGCTGCTCAACGAGC 9323  
Qy 863 TGACCGTGGCGAGCGGAGCTCCCAAGAAATCGTCTGAAACGGGCGCTGCGAGCGAG 922  
Db 9324 TGACCGGCGCGACCGGCGCGAGCGGATGCGGAGCGGCTGCTCCGACGCGCG 9383  
Qy 923 GCTCGCGCGCTCTTCTGAGGATATGTCGAGGACAGCGAGCGGCGAGCGCTGTG 982  
Db 9384 GCGTGTCTCCCGAGAGAGCTGTGACCGGCTGTGAGCGGACAGCGAGCGCGCGCG 9443  
Qy 983 ACCCATGGAATCAAGCTCTGAAATGCGGATACGCGCTCGGCGAGATGTCCGACGC 1042  
Db 9444 ACCCATGAGAGCGGCGCGCTGCTGCGCGCTCGGACGGAACGTTCCGCGACAC 9503  
Qy 1043 CGCTGTGATGAGGCTGTGAGAGACCACTTTGGCATCTGAGATATCGTGGAGATCA 1102  
Db 9504 CGCTGTGCTGTGCTGTGAGATCAATCGGAGATGCGGAGCGCGCGCGGCTGTG 9563  
Qy 1103 CTGGGCTGTGAGAGTGTCTTGTCTTCTGACAGCGGAGATTTCTGCGACCTCCAG 1162  
Db 9564 GCGGCTGTATCAAGATGCTTCAAGCGGCTGTGCGGACGCTTGTGCTCCGACCTT 9623  
Qy 1163 CGGAGCGCTGAACCCCGGATCTCATGAGGATCTTGTGCTGACCGTCAAGCGCGCC 1222  
Db 9624 CGGAGAGCGAGCCCGCATGCGAGCTGAGCTCGGCGGATGAGGCTGTGCTCACT 9683  
Qy 1223 GGAACCGTGGCGGAGCTGGAATACCGCGGACGCGGCGGAGTGTGCTGTTCGATGA 1282  
Db 9684 AGGTCCCTGTGCGAGGAGCGGCGCGCGCGCGGAGTGTTCCTTGTGCGCGTGC 9743  
Qy 1283 GCGGAGCAAGCGGCGAGTGTGCTGGAAGAGCGCGCGGCGGAGCTGCAACCGCGCG 1342  
Db 9744 GCGGAGCAAGTGTGCTGCTGTGAGAGAGCACCGCGCGCGCGCGGAGAGCGG 9803  
Qy 1343 GCGGAGCGAGC-----GGCAGAGCTGTG 1369

Db 9804 CCGGAGAGCGCCCGCGGCTCCCGCGCGCGGAGAGGCGGAGAGGCGCTGTGCGG 9863  
Qy 1370 TGCTGTGCGAAGAGACCGCGCTGAGCTGTGATGACAGGCGCGCGGCTGTGCGGACCATC 1429  
Db 9864 TGCTGTGCGAAGAGAGAGCGGCTGTGAGCTGTGAGAGCGGCGGCTGTGCGGACCATC 9923  
Qy 1430 TGAAGACCTACCTTGTGAGTGTGAGGAGATGTGAGGCTGTGAGGAGAGCGGCA 1489  
Db 9924 TCTCCGACACCCCGGCGCGCGCGGATGATGCTGCTTCTCTGCGCGGACGCGCG 9983  
Qy 1490 GCGGATGAGAGACCGGCTGTGAGGAGAGAGTGTGAGGAGAGGAGGCTGTGAGGAG 1549  
Db 9984 CAGCTTGTGACACCGCGCGGCTGTGAGTGTGAGAGAGGAGGAGGCTGTGAGGAG 16043  
Qy 1550 TGAAGCTGTGCGGCGGAGAGAGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1609  
Db 10044 TGAAGCTGTGCGGAGAGAGAGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 10100  
Qy 1610 CAGCGGAGAGCTGTGCTTCTTCAACGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669  
Db 10101 GAGAGCGGAGAGTGTGCTTCTTCAACGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 10160  
Qy 1670 GTGAGCTGTGAGATGATGATGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1729  
Db 10161 ACAGCTGTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10220  
Qy 1730 TCAACGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1789  
Db 10221 TGAAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10280  
Qy 1790 ACAGCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1849  
Db 10281 AAGCGGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10340  
Qy 1850 GCTGTGCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909  
Db 10341 GCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10400  
Qy 1910 TCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1969  
Db 10401 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10460  
Qy 1970 TGTGTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2029  
Db 10461 TGTGTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10520  
Qy 2030 TCGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2089  
Db 10521 TGGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10580  
Qy 2090 TGGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2149  
Db 10581 TGGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10640  
Qy 2150 CGATGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2209  
Db 10641 AGGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10700  
Qy 2210 GCTTCACTCAACGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2269  
Db 10701 CTTTCACTCAACGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10760  
Qy 2270 TGAAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2326  
Db 10761 TGGCTTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10820  
Qy 2327 CAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2386  
Db 10821 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10880  
Qy 2387 CGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2446  
Db 10881 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10940





Dh 3296409 GAGCGCGCGGTTGAGCGGAGGAAAGCACCGGGGGCCGCGCCCTGATGTGCGGTGG 3296350  
Qy 1370 TGCTGTGCGCAAGAGACCGCGCTCAGCCCTGATGACAGAGCGCGCGGCTGCGCAATC 1429  
Dh 3296494 TGATTTGGGCGAGATCGCTGAGCGCTTGAACGGCGAGCGCGGTGCACTTATGCGCCACG 3296290  
Qy 1430 TGAAGACCTACCTTTCGAGTGTCTGAGCGATGTGGCGTTCAGTCTGGCGACGACGCA 1489  
Dh 3296289 TGCAGGCGCAACCCAGGGGTGATCCGATGATGTGGGTGTCTGTTGGCCAGT---CGCT 3296233  
Qy 1490 GCGGATGAGACACCGGCTCGCGGTGGCGGACGTCGAGGGAGGGGTTGGGGCAGCC 1549  
Dh 3296232 CGGTGTTGAGCACCGAGCGGTGGTGTGCGCGCAAGCCGTGACCACTGATTCGCGGC 3296173  
Qy 1550 TGAACGCTGCGCGCCAGGAGACAGAGTGCAGCGGAGCGGTGCGGAGTATGCGCGATTCCT 1609  
Dh 3296172 TGGCTGGGCTGCGCGCGCGGCGAGCCGGGTGCGCGGTGGCTGCTGACGACAG---GCT 3296116  
Qy 1610 CACGCGGCAAGCTGCGCTTCTCTTCAACCGACAGGGGCGCAGACGCTGGGCAATGGGCC 1669  
Dh 3296115 CGGTGGCGAAGACGTTGGTCTGTTTCTTGGGAGGGGCGCAGCGCATGGGATGGGCC 3296056  
Qy 1670 GTGGGCTGTACGATGATGTGCTCGCGCTTCCGCGAGGCGTTGACCTGTGCTGAGGCTGT 1729  
Dh 3296055 GCGAGTTGTACGGGAGTGTGCTCGTTTGGCGAGGCAATGATGCGGTGCGCAGCAGT 3295996  
Qy 1730 TCAACGAGAGCTGACCGCGCGCGCTCGGAGGATGTGGGGCGGAACCGGCGACGCTCG 1789  
Dh 3295995 TGAACCGGCACTTCGCGGTGCGCGCTGCGCGAGCTTATTTGGG-----GTGCGG 3295948  
Qy 1790 ACGCGCGCTGCTGACACGACACCTTCAACCGCGCGCGCTGTTCACTTTCGAAATATG 1849  
Dh 3295947 ATGCGGATTTGCTTGAAGACGCAATTTGCTCAAGCCGCGCTTGTGGCGGTGAGGTCG 3295888  
Qy 1850 CGCTCGCGCGCTGTGGCGGTGCTGGGGTGTAGAGCCGAGATTGTGCGCGCCATAGCA 1909  
Dh 3295887 CATGCTTCGCGGTGTTCGGGATTTGGGGTGTGCTTCGCGACTTCGTCAATGGGTCACTCCG 3295828  
Qy 1910 TCGGTGAGCTGTGAGCTGCTGCGGTGCGGCGGCGGTCTGCTGAGAGCGCGGTTTC 1969  
Dh 3295827 TTGAGAGAGCTTGGCGCGCGCGCGACCGCGCGGTGTGTGATCTTGTGCGGACCGCGGATGC 3295768  
Qy 1970 TGTGTGCTGCGCGCGCGCGCTGTATGACAGCGCTGCGCGCGCGCGCGATGTGTGA 2029  
Dh 3295767 TGTGTGTGGCGCGCGCGCGGTGTATGACAGCGCTGCGCGCGCGCGCGATGTGTGGCG 3295708  
Qy 2030 TCGAGCGCGCGAGCGGATGTGCTGCTGCTGCGCGCGCGCGACGACGCTGCTGTGA 2089  
Dh 3295707 TGGCTGCGAGTGAAGACGAGGT-----GAGCGCGCTGCTGGGTGAGGGTGTGGGGA 3295657  
Qy 2090 TCGCGCGGCTGAAGCTTCGCGGACAGAGGTGATTCGCGCGCGCGCGGCAACCGGTGATG 2149  
Dh 3295656 TCGCTGCGATTAACGCGCGCGCAATTCGGTGTGATCTTCGGTGCGCGCGCGCGCAATG 3295597  
Qy 2150 CGATCGCGCGCGAGTGGCGCGCGCGCGCGCGCGCAACCAAGCGCTCCACGCTTCGATG 2209  
Dh 3295596 CGATTGGGAGATCGGTTGCGCGCGAGAGGTTCGCGCGGTGCAACAGTTGGCGGTTCGATG 3295537  
Qy 2210 CGTTTCACTACCGCTCAATGCGCGCGCGATGTGAGAGCGTTGCGCGGTGTGGCGGATCG 2269  
Dh 3295536 CGTTTCACTTCGCGCTGTATGAGCGATGCTGAGAGATTGCGCGGTGTGCGCGCGCGG 3295477  
Qy 2270 TGAAGCTACCGCGCGCGCGCTGATGCTCTGCTGACAGCAATCTGAGGGGAGAGGCTTGCACAG 2329  
Dh 3295476 TGCAGGCAACGAGACCCCGAGCTTGGCTGTGATGATGCAACGTACGAGGAGTTGGCGCGCC 3295417  
Qy 2330 ACGAGGTGAGCTTCGCGCGGCTATTTGGGTGCGCGCACGCGCGAGAGGTGTGCGTTTCGCG 2389  
Dh 3295416 CTGATTTTCGGGTTCGCGCGAGTACGGGTGAGCAAGTTTCGTTCGCGCGGTTCGCTTCGCG 3295357  
Qy 2390 ATGAGTGAAGGCGCTGACCGCGCGGTGCGCGGACCTTGTGAGGTGCGTTCGAAT 2449

Dh 3295356 ACAGTGCGGCTATTGACAGACCTTGGGGGAGACCACTTCATGAGAGCGCGCGGAAA 3295297  
Qy 2450 GAGAGCTGCTCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2509  
Dh 3295296 GTGTTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3295237  
Qy 2510 GTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2569  
Dh 3295236 TGTGCGGCAAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3295177  
Qy 2570 CCGTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2626  
Dh 3295176 CCACGGGTGTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3295117  
Qy 2627 CGTGTCCAGATACCTTGTGCGAGCGCGCTACTGATGACACGAAAGCCGACGACG 2686  
Dh 3295116 AGCTGCCAGATATGCTTTCAGGAGAGCGCGTTTGGAGAGCGCGCGCGCGATGGGC 3295057  
Qy 2687 CCGCGGTGCGACCGCGCTGCTTCGAGAGCGGCTACGACGAGTTCGAGAGGCGGCGG 2746  
Dh 3295056 CCGCGATGCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3294997  
Qy 2747 CGGTGCGCGCGG 2758  
Dh 3294996 TCGAGCGCGCGG 3294985

Search completed: October 4, 2003, 03:39:32  
Job time : 275.999 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 3, 2003, 11:29:26 ; Search time 1043.88 Seconds  
(without alignment)

11031.749 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_7610\_11875

Perfect score: 4266

Sequence: 1 gtggcgatgcgtccatcga.....gggaatcattgcccataga 4266

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4266	100.0	68750	21	AAZ55887
2	4187.6	98.2	71989	21	AAA29349
3	817.2	19.2	67459	25	ABO77491
4	814.2	19.1	3978	21	AAZ55785
5	814.2	19.1	12349	21	AAZ55840
6	814.2	19.1	18331	21	AAZ55857
7	805	18.9	27541	22	AAD17185
8	805	18.9	125401	22	AAD17186

9	803.2	18.8	33529	23	AAZ17367
10	766.6	18.0	20394	22	AAZ24892
11	755.2	17.7	43280	18	AAZ80413
12	745.8	17.5	53789	19	AAV21187
13	742	17.4	15872	21	AAZ87283
14	740.8	17.4	441529	22	AA196682
15	735.6	17.2	15872	18	AAZ8715
16	732	17.0	4403765	22	AA196683
17	726.2	17.0	65140	22	AAD17184
18	725	17.0	49377	19	AAV05287
19	701.2	16.4	16767	22	AAZ88339
20	701.2	16.4	50000	22	AAZ88313
21	679.2	15.9	80161	20	AAZ21501
22	679.2	15.9	14775	22	AAZ88336
23	679.2	15.9	50000	22	AAZ88316
24	677.2	15.9	75236	25	AAV5557
25	675.2	15.8	44377	18	AAZ78508
26	675.2	15.8	44377	18	AAZ80414
27	671.8	15.7	28598	17	AAZ06769
28	671.8	15.7	28958	18	AAZ89956
29	671.8	15.7	28958	21	AAZ75299
30	671	15.7	9513	22	AAZ88337
31	663.2	15.5	29879	14	AAQ46806
32	662.2	15.5	13987	18	AAZ80415
33	649.4	15.2	47981	22	AAZ30757
34	639	15.0	6459	22	AAZ88336
35	633.2	14.8	5088	22	AAZ88336
36	633.2	14.8	34071	22	AAZ88336
37	633.2	14.8	42717	22	AAZ88336
38	633	14.8	77536	21	AAZ14651
39	628.8	14.7	4403765	22	AAZ88336
40	628.8	14.7	441529	22	AAZ88336
41	627.6	14.7	13842	21	AAZ88336
42	627.6	14.7	36778	21	AAZ87318
43	627.6	14.7	37948	21	AAZ87285
44	627.6	14.7	38506	21	AAZ15633
45	627.6	14.7	38506	21	AAZ56001

## ALIGNMENTS

AAZ55887	standard; DNA; 68750 BP.
AAZ55887	10-APR-2000 (first entry)
AAZ55887	Sorangium cellulosum 68.75 kb contig.
AAZ55887	Epithelone biosynthesis; type I polyketide synthase; taxol substructure; anticancer; ds.
AAZ55887	Sorangium cellulosum.
AAZ55887	Location/Qualifiers
AAZ55887	1..1826
AAZ55887	/tag= a
AAZ55887	/product= "partial Orf 1 protein (AAZ55887)"
AAZ55887	/note= "No initiation codon given in the specification"
AAZ55887	complement (1900..3171)
AAZ55887	/tag= b
AAZ55887	/product= "Orf 2 protein (AAZ55887)"
AAZ55887	3415..5556
AAZ55887	/tag= c
AAZ55887	/product= "Orf 3 protein (AAZ55887)"
AAZ55887	complement (5612..5992)
AAZ55887	/tag= d
AAZ55887	/product= "Orf 4 protein (AAZ55887)"
AAZ55887	6226..6675

DNA sequence of S.  
Pimaricin biosynth  
Tylactone synthase  
Amycolatopsis medi  
S. venezuelae vep  
Mycobacterium tube  
Streptomyces venez  
Mycobacterium tube  
Streptomyces nours  
The soraphen biosy  
S. spinosa DNA fra  
S. spinosa DNA fra  
DNA fragment of Sa  
S. spinosa DNA fra  
S. spinosa DNA fra  
Saccharopolyspora  
Platenolide syntha  
Sorangium cellulos  
DNA sequence of So  
S. spinosa DNA fra  
eryl region of S.  
Hybrid smg/tylg O  
Micromonospora meg  
S. spinosa DNA fra  
Nucleotide sequenc  
Nucleotide sequenc  
Nucleotide sequenc  
Nucleotide sequenc  
Mycobacterium tube  
Mycobacterium tube  
S. venezuelae macr  
S. venezuelae pik  
S. venezuelae pik  
Nucleotide sequenc  
Recombinant cosmid

```

FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX
XX      W09966028-A2.
XX
XX      23-DEC-1999.
XX
XX      16-JUN-1999; 99WO-EP04171.
XX
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
XX      05-FEB-1999; 99US-0118906.
XX

```

```

PA      (NOVS ) NOVARTIS AG.
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERN GES MBH.
XX
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D,
XX      WPI: 2000-097741/08.
XX      P-P8DB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
XX      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
XX      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
XX      AAV58592, AAV58593, AAV58594.
XX
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy
XX
XX      Claim 14; Page 87-104; 174pp; English.
XX
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-158578) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiply drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
XX
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX      Query Match 100.0%; Score 4266; DB 21; Length 68750;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      1 GTGGCGGATCGTCCATCGAGGCGCGAGCGGAGATCGATTGCGATGCGAGCGAGT 60
XX      7610 GTGGCGGATCGTCCATCGAGGCGCGAGCGGAGATCGATTGCGATGCGAGCGAGT 7669
XX      61 TCCCGTCTGCCCGATGCGGCGTGAATCGATGAGCGGGTTCTGGACGCTCTCGAGGGCTCG 120
XX      7670 TCCCGTCTGCCCGATGCGGCGTGAATCGATGAGCGGGTTCTGGACGCTCTCGAGGGCTCG 7729
XX      121 CGCGACATCGTGGGCGAGTCCCCCGAGCGCTGGATGACGACGCGTGTGTTATCCC 180
XX      7730 CGCGACATCGTGGGCGAGTCCCCCGAGCGCTGGATGACGACGCGTGTGTTATCCC 7789
XX      181 GACCCGATGCGCGGAGGAGAGCGCGCTTACGCGCGATCTTTCTGAGCAGCTACCC 240
XX      7790 GACCCGATGCGCGGAGGAGAGCGCGCTTACGCGCGATCTTTCTGAGCAGCTACCC 7849
XX      241 TCCCTGAGCGCTCTCTTCTTGGCATCTCGCTCGGAGCGCTGCGGATGAGACCTTGA 300
XX      7850 TCCCTGAGCGCTCTCTTCTTGGCATCTCGCTCGGAGCGCTGCGGATGAGACCTTGA 7909
XX      301 CATGACCTCTTCTGAGAGTGTGCTGGAGGCGCTGGAGGAGGCGCGGATGAGCTCTCATCG 360
XX      7910 CATGACCTCTTCTGAGAGTGTGCTGGAGGCGCTGGAGGAGGCGCGGATGAGCTCTCATCG 7969
XX      361 GCGCTGTGCTGAGGAGGAGGAGTGTTCATCGGATCGGCGCTGCGGATATGAGGCC 420
XX

```

Db 7970 GCGCTGTCGATGACGAAACGGAGTGTTCATCGGAAATCGGCCCTCCGAATATAGACC 8029  
Qy 421 GCGCTGCCGAAGCAGACGGCGCTCCGACAGATCGACGCTCATGGCGGGCTGGAGACGATG 480  
Db 8030 GCGCTGCCGAAGCAGACGGCGCTCCGACAGATCGACGCTCATGGCGGGCTGGAGACGATG 8089  
Qy 481 CCCAGCGTCGAGAGCGGGCGGAATCTGTATGCGCTCGGGCTGCGAGGGCCGTGTGTCGCG 540  
Db 8090 CCCAGCGTCGAGAGCGGGCGGAATCTGTATGCGCTCGGGCTGCGAGGGCCGTGTGTCGCG 8149  
Qy 541 GTGATAGCGGCTTATTCGTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Db 8150 GTGATAGCGGCTTATTCGTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8209  
Qy 601 TCCGGGAATGCTCCAGCGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
Db 8210 TCCGGGAATGCTCCAGCGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8269  
Qy 661 CTCTGTGCTCTTCGAGAGACCGGGCGCTGCGCAGAGAGCGTCTGCTGCAAGCATTTTCG 720  
Db 8270 CTCTGTGCTCTTCGAGAGACCGGGCGCTGCGCAGAGAGCGTCTGCTGCAAGCATTTTCG 8329  
Qy 721 GCGAGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
Db 8330 GCGAGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8389  
Qy 781 AGTGAAGCGCGCGCGAGAGCGGCGATGAGATTTGGGAGTGTGCTGAGAGATTCGCGCATTCAT 840  
Db 8390 AGTGAAGCGCGCGCGAGAGCGGCGATGAGATTTGGGAGTGTGCTGAGAGATTCGCGCATTCAT 8449  
Qy 841 CACGACGCTGCGAGAGCGGCTGCTGACCGTGCAGAACGCGAGAGCTCCCAAGAAATCGTGTG 900  
Db 8450 CACGACGCTGCGAGAGCGGCTGCTGACCGTGCAGAACGCGAGAGCTCCCAAGAAATCGTGTG 8509  
Qy 901 AAACGGGCGCTGCGCGAGAGCGAGGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
Db 8510 AAACGGGCGCTGCGCGAGAGCGAGGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8569  
Qy 961 GGCACGGGCGAGAGCGGCTGCTGAGACCCCATTCGAAATTCGAAATTCGAAATTCGAAATTCG 1020  
Db 8570 GGCACGGGCGAGAGCGGCTGCTGAGACCCCATTCGAAATTCGAAATTCGAAATTCGAAATTCG 8629  
Qy 1021 CTCTGCGGAGATGCTGCGACCGCGCTGCTGATCGGGTCTGCTGAGAACCAACCTTGCGCAT 1080  
Db 8630 CTCTGCGGAGATGCTGCGACCGCGCTGCTGATCGGGTCTGCTGAGAACCAACCTTGCGCAT 8689  
Qy 1081 CTCTGAGTATGCTGCGGGGATCACTGGGCTGCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTG 1140  
Db 8690 CTCTGAGTATGCTGCGGGGATCACTGGGCTGCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTG 8749  
Qy 1141 CAGATTCCTGCGACCTTCACGCGGAGCGGCTGAAACCCCGGATTCATGAGGCTGATCTT 1200  
Db 8750 CAGATTCCTGCGACCTTCACGCGGAGCGGCTGAAACCCCGGATTCATGAGGCTGATCTT 8809  
Qy 1201 CGGCTGACCGTACCGCGCGCGCGGACACCGTGGCGGACTGGATATACCCCGGACGCGCG 1260  
Db 8810 CGGCTGACCGTACCGCGCGCGCGGACACCGTGGCGGACTGGATATACCCCGGACGCGCG 8869  
Qy 1261 GGGGTGAGCTGCTTGGCATGAGCGGAGCAACCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
Db 8870 GGGGTGAGCTGCTTGGCATGAGCGGAGCAACCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8929  
Qy 1321 GCGGAGAGTGCACACCGCGCGCGCGGAGCGGACCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
Db 8930 GCGGAGAGTGCACACCGCGCGCGCGGAGCGGACCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTG 8989  
Qy 1381 AGGACCGGCTACGCTTGGATGCAAGGCGGCGCGGCTGCGGACCATCTGAGAGCTTAC 1440  
Db 8990 AGGACCGGCTACGCTTGGATGCAAGGCGGCGCGGCTGCGGACCATCTGAGAGCTTAC 9049  
Qy 1441 CTTTGCAGTGTCTTGGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500

Db 9050 CTTTGCAGTGTCTTGGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9109  
Qy 1501 CACCGGCTCGCGGTGCGCGGACGCTGCGAGAGGAGGCTGCGGGACGCTTGAACGCTGCG 1560  
Db 9110 CACCGGCTCGCGGTGCGCGGACGCTGCGAGAGGAGGCTGCGGGACGCTTGAACGCTGCG 9169  
Qy 1561 GCGAGAGGACAGAGTGCAGCGGCTGCGGCTGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTG 1620  
Db 9170 GCGAGAGGACAGAGTGCAGCGGCTGCGGCTGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTG 9229  
Qy 1621 CTGCGCTTCTCTTCAACCGGACAGGGGCGGAGACGCTGCGGACATGCGGCTGCTGCTGCTGCTG 1680  
Db 9230 CTGCGCTTCTCTTCAACCGGACAGGGGCGGAGACGCTGCGGACATGCGGCTGCTGCTGCTGCTG 9289  
Qy 1681 GATGATAGTCCGAGCTTCCGAGAGGCTTCCAGCTGTCGCTGAGAGGCTTCAACAGAGAG 1740  
Db 9290 GATGATAGTCCGAGCTTCCGAGAGGCTTCCAGCTGTCGCTGAGAGGCTTCAACAGAGAG 9349  
Qy 1741 CTGACCGGCGGCTTCCGAGAGTATGAGGCGGACCGGCGAGGCTGACGCGCGCTG 1800  
Db 9350 CTGACCGGCGGCTTCCGAGAGTATGAGGCGGACCGGCGAGGCTGACGCGCGCTG 9409  
Qy 1801 CTGACCGAGAGAGCTTCAACCGGAGCGGCTGCTTCACTTCAATATGCGCTGCGCGCG 1860  
Db 9410 CTGACCGAGAGAGCTTCAACCGGAGCGGCTGCTTCACTTCAATATGCGCTGCGCGCG 9469  
Qy 1861 CTGTGCGGCTGCTGAGGAGTATGAGCGGAGGTTGCTGCGCGGACATAGCATGCTGAGCTG 1920  
Db 9470 CTGTGCGGCTGCTGAGGAGTATGAGCGGAGGTTGCTGCGCGGACATAGCATGCTGAGCTG 9529  
Qy 1921 GTGCTGCTGCTGCTGAGGCGGCTTCTGCTTGAAGACGCGGCTTCTGCTGCTGCTGCTGCTG 1980  
Db 9530 GTGCTGCTGCTGCTGAGGCGGCTTCTGCTTGAAGACGCGGCTTCTGCTGCTGCTGCTGCTG 9589  
Qy 1981 CCGCGGCGCTGATGACAGGCGCTGCGGCGCGGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTG 2040  
Db 9590 CCGCGGCGCTGATGACAGGCGCTGCGGCGCGGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTG 9649  
Qy 2041 GAGGCGGAGTGTGCTGCTGCGGCTGCGGCGCGGACGAGGCTGCTGCTGCTGCTGCTGCTGCTG 2100  
Db 9650 GAGGCGGAGTGTGCTGCTGCGGCTGCGGCGCGGACGAGGCTGCTGCTGCTGCTGCTGCTGCTG 9709  
Qy 2101 AACGCTCCGAGCAAGTGTGCTGCTGCGGCGCGGCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTG 2160  
Db 9710 AACGCTCCGAGCAAGTGTGCTGCTGCGGCGCGGCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTG 9769  
Qy 2161 GCGATGCGCGCGCGCGGCGGCGGAGCAACAGGCGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTG 2220  
Db 9770 GCGATGCGCGCGCGCGGCGGCGGAGCAACAGGCGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTG 9829  
Qy 2221 CCGCTCATGCGCGGAGTGTGAGAGGCTTGGGCGGCTGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTG 2280  
Db 9830 CCGCTCATGCGCGGAGTGTGAGAGGCTTGGGCGGCTGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTG 9889  
Qy 2281 CCGGCTGCTGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340  
Db 9890 CCGGCTGCTGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9949  
Qy 2341 TCGCGGCGCTATTTGGGCTGCGGACGCGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400  
Db 9950 TCGCGGCGCTATTTGGGCTGCGGACGCGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10009  
Qy 2401 GCGGTCGACCGGCGCGGCGGAGGCTTGTGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2460  
Db 10010 GCGGTCGACCGGCGCGGCGGAGGCTTGTGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10069  
Qy 2461 GCGCTG 2520  
Db 10070 GCGCTG 10129  
Qy 2521 GCGGCTGACAGGCGGCGGAGCGGCTGCTGAGAGGCTTGGCGGCTTGTGGGCTGCTGCTGCTGCTGCTG 2580  
Db 10130 GCGGCTGACAGGCGGCGGAGCGGCTGCTGAGAGGCTTGGCGGCTTGTGGGCTGCTGCTGCTGCTGCTGCTGCTG 10189

QY	2581	CTGGTCTCTGAGGAGCGGCGCTCTTCCCTTCAGAGGGGGAGCGGGGTGCGCTGCACCAAGTAC	2640
Db	10190	CTGGTCTCTGAGGAGCGGCGCTCTTCCCTTCAGAGGGGGAGCGGGGTGCGCTGCACCAAGTAC	10249
QY	2641	CTTTGGCAGCGCGACGCGCTACTGGATGTGACACGAAAGCCGACGCGCGCGCTGGCGAC	2700
Db	10250	CTTTGGCAGCGCGACGCGCTACTGGATGTGACACGAAAGCCGACGCGCGCGCGCTGGCGAC	10309
QY	2701	CGCGGTGCTCGGAGAGCGGGTCAAGACGAGGTGCAGGAGGGGGGCGCGGTGCGCGCGAC	2760
Db	10310	CGCGGTGCTCGGAGAGCGGGTCAAGACGAGGTGCAGGAGGGGGGCGCGGTGCGCGCGAC	10365
QY	2761	GACCGGCGCAGCGCTCGGCTTCGACCATCGCGCGCGAGAGCGGACGCGCGAGGAAGTTC	2820
Db	10370	GACCGGCGCAGCGCTCGGCTTCGACCATCGCGCGCGAGAGCGGACGCGCGAGGAAGTTC	10429
QY	2821	GAGGCGCGCGCGGACCGTCCGCTTCCGGCTGAGATGCATGACGCAAGCGCTGTGATTCAC	2880
Db	10430	GAGGCGCGCGCGGACCGTCCGCTTCCGGCTGAGATGCATGACGCAAGCGCTGTGATTCAC	10489
QY	2881	CTCGTGTCTTGGGATCAGAGAGCGGCGCGCCCTGTGTGTGGGCGAGGTTCAGATCGCGTTC	2940
Db	10490	CTCGTGTCTTGGGATCAGAGAGCGGCGCGCCCTGTGTGTGGGCGAGGTTCAGATCGCGTTC	10549
QY	2941	GACCGCGCGGCGCTCAGCTTCAATGATGTCCAGTCTGCGCGTGGGCAATGTGTCCGACGAC	3000
Db	10550	GACCGCGCGGCGCTCAGCTTCAATGATGTCCAGTCTGCGCGTGGGCAATGTGTCCGACGAC	10609
QY	3001	CTGCGCGGGAAGGCCCAACCCCTCCGCTGTCTGTGAGAGCGAGTGC CGCGCGGCGCAATCGTAC	3060
Db	10610	CTGCGCGGGAAGGCCCAACCCCTCCGCTGTCTGTGAGAGCGAGTGC CGCGCGGCGCAATCGTAC	10665
QY	3061	GCCGTGGGCGAGGGCGTGAAAGGCTGTGTGTGGGCCCAACCGGTCAATCGCCCTTTGGGCG	3120
Db	10670	GCCGTGGGCGAGGGCGTGAAAGGCTGTGTGTGGGCCCAACCGGTCAATCGCCCTTTGGGCG	10729
QY	3121	GGACCGTTTGGTACCAACGTCAACAAGTCCGCTGGTGTGTGCTCGGCTTCAGGGG	3180
Db	10730	GGACCGTTTGGTACCAACGTCAACAAGTCCGCTGGTGTGTGCTCGGCTTCAGGGG	10789
QY	3181	CTCTCGGCGATTCAGAGCGGCGCGCCATCCCGCTGCGGTACCTGACGCGCATGTATCGCGCTC	3240
Db	10790	CTCTCGGCGATTCAGAGCGGCGCGCCATCCCGCTGCGGTACCTGACGCGCATGTATCGCGCTC	10849
QY	3241	GACGGAATATGCGCGCCCTTACAGCGCGGGGAGCGGGTGTGATTCATGTGCGCGACCGCGCGG	3300
Db	10850	GACGGAATATGCGCGCCCTTACAGCGCGGGGAGCGGGTGTGATTCATGTGCGCGACCGCGCGG	10909
QY	3301	GTCGGTCTCGCGGCGGTGACAGTGGGCGCACAGTGGGAGCCGAGGTGCATGCGAGCGGCG	3360
Db	10910	GTCGGTCTCGCGGCGGTGACAGTGGGCGCACAGTGGGAGCCGAGGTGCATGCGAGCGGCG	10965
QY	3361	GGCACGCGCCGAGAAACGCGCTTACCTTGAGTGTGTGGGCTGTGCGGTATGTGACCAATTCC	3420
Db	10970	GGCACGCGCCGAGAAACGCGCTTACCTTGAGTGTGTGGGCTGTGCGGTATGTGACCAATTCC	11029
QY	3421	CGCTTCGAGACCGGTTCTGTCTGCGACAGTGTGCGCGCTGTGACCGGCGCGGAGGATGACGTTTC	3480
Db	11030	CGCTTCGAGACCGGTTCTGTCTGCGACAGTGTGCGCGCTGTGACCGGCGCGGAGGATGACGTTTC	11089
QY	3481	GTCGCTCAACTGCTCTCGGGGCGAGCTATGACACGAGGTTTCAATCTCTCGCGCATGCGAC	3540
Db	11090	GTCGCTCAACTGCTCTCGGGGCGAGCTATGACACGAGGTTTCAATCTCTCGCGCATGCGAC	11149
QY	3541	GGC CGGATTGTGAGCTTCGCGACGCGCATGTTTACGCGGATTAACAAGTCTCGGCGTTCGCG	3600
Db	11150	GGC CGGATTGTGAGCTTCGCGACGCGCATGTTTACGCGGATTAACAAGTCTCGGCGTTCGCG	11209
QY	3601	CGGTTCCGCGCAATCTCTCTTCTGCGTGGTGGATCTCGCGGGAGATGATGCTTCGAGCGG	3660
Db	11210	CGGTTCCGCGCAATCTCTCTTCTGCGTGGTGGATCTCGCGGGAGATGATGCTTCGAGCGG	11265

Oy		3661	CCGGCGGAGGATCCGTGGGCTCTTGGAGAAGACTCTGGGCTCATCGCGCAGACCGTGTC	3720			
Db		11270	CCGGCGGAGGATCCGTGGGCTCTTGGAGAAGACTCTGGGCTCATCGCGCAGACCGTGTC	11329			
Oy		3721	ACCCTCCCCCAATCGGAGCGCTCCCAGATGCCGTGTGCAGATGCCGTTCCGAGCATG	3780			
Db		11330	ACCCTCCCCCAATCGGAGCGCTCCCAGATGCCGTGTGCAGATGCCGTTCCGAGCATG	11389			
Oy		3781	GCCGAGGCGGAGCATCTTGGGAAGCTCGTAATCACGCTGGGTGACCCGGAGGATCCAGATC	3840			
Db		11390	GCCGAGGCGGAGCATCTTGGGAAGCTCGTAATCACGCTGGGTGACCCGGAGGATCCAGATC	11449			
Oy		3841	CGATTTCACAACCCAGCAGGCGCGGACCCGTCACACGGGGATCGGGAACCTGCTCGACAGG	3900			
Db		11450	CGATTTCACAACCCAGCAGGCGCGGACCCGTCACACGGGGATCGGGAACCTGCTCGACAGG	11509			
Oy		3901	CTCGCGTCAAGCTCGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCTCCGTAGCAG	3960			
Db		11510	CTCGCGTCAAGCTCGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCTCCGTAGCAG	11569			
Oy		3961	GTCCTGCAGGTGCTGCGCACGCCCCAATCAGGTGGCGCGGAGCGCTGTTCAACCGC	4020			
Db		11570	GTCCTGCAGGTGCTGCGCACGCCCCAATCAGGTGGCGCGGAGCGCTGTTCAACCGC	11629			
Oy		4021	CTCGGCAATGGAACTCGCTCATAGGCGCGGAGCTGCGCAATCGATCGAGGCGAGCTCAAG	4080			
Db		11630	CTCGGCAATGGAACTCGCTCATAGGCGCGGAGCTGCGCAATCGATCGAGGCGAGCTCAAG	11689			
Oy		4081	CTGAACTGTTCGACGACAGTTCCTGTCCAGGTCGCCCAATATGCGCTTGTGGCCGAAAAC	4140			
Db		11690	CTGAACTGTTCGACGACAGTTCCTGTCCAGGTCGCCCAATATGCGCTTGTGGCCGAAAAC	11749			
Oy		4141	CTGTTGATGCTCTTCGCGACAGCTCTCTCTTTGAGCGGGTGGCGCGGAGAACCTACGG	4200			
Db		11750	CTGTTGATGCTCTTCGCGACAGCTCTCTCTTTGAGCGGGTGGCGCGGAGAACCTACGG	11809			
Oy		4201	GCGAGCGCTGAAAAACGACTTCGTCTCATGGGCGCGACATCAAACCTGGGAAATCAATTGCC	4260			
Db		11810	GCGAGCGCTGAAAAACGACTTCGTCTCATGGGCGCGACATCAAACCTGGGAAATCAATTGCC	11869			
Oy		4261	CTATGA 4266 				
Db		11870	CTATGA 11875				
<hr/>							
RESULT 2							
ID	AAA29349						
XX	AAA29349 standard; DNA; 71989 BP.						
XX							
AC	AAA29349;						
XX							
DT	12-SBP-2000 (first entry)						
DE	Sorangium cellulosum epochilone polyketide synthase operon genomic DNA.						
XX							
KW	Epochilone; polyketide synthase; epoa; epob; epoc; epod; epof; epog; epol; epok; p450 epoxidase; ORFa; ORFb; promoter; enhancer; anti-fungal; tubulin polymerization assay; anti-tumor; cytostatic; ds.						
XX							
OS	Sorangium cellulosum.						
XX							
Key	Location/Qualifiers						
FT	CDS	3..992	/label= a				
FT			/*cag= a				
FT			/product= transposase				
FT			/note= "not part of the PKS"				
FT	CDS	989..1501	/*cag= b				
FT			/label= ORF B				
FT			/product= transposase				
FT			/note= "not part of the PKS"				
FT			1998..6263				
FT	CDS						



```
FT      /*tag= c
FT      /label= epoc_gene
FT      /note= "encodes the loading domain"
FT      2031..3548
FT      /*tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /*tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /*tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /*tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /*tag= h
FT      /label= epoc_gene
FT      /note= "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /*tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /*tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /*tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /*tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /*tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /*tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /*tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /*tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /*tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /*tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /*tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /*tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /*tag= w

FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /*tag= x
FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /*tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /*tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /*tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /*tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /*tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /*tag= ad
FT      /label= epod_gene
FT      /note= "encodes modules '3-6"
FT      misc_RNA
FT      16425..17606
FT      /*tag= ae
FT      /note= "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /*tag= af
FT      /note= "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /*tag= ag
FT      /note= "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /*tag= ah
FT      /note= "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /*tag= ai
FT      /note= "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /*tag= aj
FT      /note= "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /*tag= ak
FT      /note= "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /*tag= al
FT      /note= "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /*tag= am
FT      /note= "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /*tag= an
FT      /note= "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /*tag= ao
FT      /note= "encodes DHS"
FT      misc_RNA
FT      29433..30287
FT      /*tag= ap
FT      /note= "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /*tag= aq
FT      /note= "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /*tag= ar
FT      /note= "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /*tag= as
FT      /note= "encodes KS6"
FT      misc_RNA
FT      33018..34067
```



Db 3678 GATGATAGCCCGCTTCGCGAGGCGGTTCAGCTGTGCTGAGGCTGTTCACACAGAG 3737  
Qy 1741 CTCGACCGGCGCTCCGAGAGGTGATGTGGCCGAACCGGACGAGGTGACACCGCGCTG 1800  
Db 3738 CTCGACCGGCGCTCCGAGAGGTGATGTGGCCGAACCGGACGAGGTGACACCGCGCTG 3797  
Qy 1801 CTCGACGAGACGCTTCACCCAGCCGCGCTGTTCACCTTCGAAATATGCGCTGCGCG 1860  
Db 3798 CTCGACGAGACGCTTCACCCAGCCGCGCTGTTCACCTTCGAAATATGCGCTGCGCG 3857  
Qy 1861 CTGTGCGGCTGTGTGGGGGTGTAGAGCCGGAAGTTGTGCGCGGACATAGCATGTGTAGCTG 1920  
Db 3858 CTGTGCGGCTGTGTGGGGGTGTAGAGCCGGAAGTTGTGCGCGGACATAGCATGTGTAGCTG 3917  
Qy 1921 GTGGCTGCTGCTGTGGCGGCGCTGTTCGCTGTAGAGACGCGGTTCTCGGTGCTGCG 1980  
Db 3918 GTGGCTGCTGCTGTGGCGGCGCTGTTCGCTGTAGAGACGCGGTTCTCGGTGCTGCG 3977  
Qy 1981 GCGGCGCGCTGTATGACAGCGCTGCGCGCGCGCGCGGCGATGTGTCATCGAGCGCG 2040  
Db 3978 GCGGCGCGCTGTATGACAGCGCTGCGCGCGCGCGCGGCGATGTGTCATCGAGCGCG 4037  
Qy 2041 GAGGCGCATGTGTGCTGTGCGGTGCGCGCGACGAGCGTGTGTCATCGAGCGCGCTC 2100  
Db 4038 GAGGCGCATGTGTGCTGTGCGGTGCGCGCGACGAGCGTGTGTCATCGAGCGCGCTC 4097  
Qy 2101 AACGCTCGGACCAAGGTGTGATGCGGGGCGCGCGGCAACCGGTGACATGCGAGCGCG 2160  
Db 4098 AACGCTCGGACCAAGGTGTGATGCGGGGCGCGCGGCAACCGGTGACATGCGAGCGCG 4157  
Qy 2161 GCGATGCGCGCGCGCGCGCGCGGCAACCAAGGCGCTCCAGTCTCGCATGCTTCACCTCA 2220  
Db 4158 GCGATGCGCGCGCGCGCGCGGCAACCAAGGCGCTCCAGTCTCGCATGCTTCACCTCA 4217  
Qy 2221 CCGCTCATGCGCGCGCGCGCGGCTGTGGCGCTGTGCGGACGAGTGTGCTGCTGCTGCTG 2280  
Db 4218 CCGCTCATGCGCGCGCGCGCGGCTGTGGCGCTGTGCGGACGAGTGTGCTGCTGCTGCTG 4277  
Qy 2281 CCGCGGTGATGCTGTGTGACGAATGAGCGGGGAGCTGTGACGAGAGGTGAGC 2340  
Db 4278 CCGCGGTGATGCTGTGTGACGAATGAGCGGGGAGCTGTGACGAGAGGTGAGC 4337  
Qy 2341 TCGCGGCGCTATGTGGTGTGCGGACGCGGAGAGGTGTGCGCTTGTGCGGATGTAGTGAAG 2400  
Db 4338 TCGCGGCGCTATGTGGTGTGCGGACGCGGAGAGGTGTGCGCTTGTGCGGATGTAGTGAAG 4397  
Qy 2401 GCGCTGACGCGCGCGGTGTGCGGACCTTTCGTGAGTGTGCGGAAATCGACGCTGCTC 2460  
Db 4398 GCGCTGACGCGCGCGGTGTGCGGACCTTTCGTGAGTGTGCGGAAATCGACGCTGCTC 4457  
Qy 2461 GCGCGGTGCTGCTGTGTGACGGAATGAGCGGGGCGCGGCTGTGCAATGCTGCGGCGCT 2520  
Db 4458 GCGCGGTGCTGCTGTGTGACGGAATGAGCGGGGCGCGGCTGTGCAATGCTGCGGCGCT 4517  
Qy 2521 GCGCGGTGACGAGCGCGGACGCTGTGAGGCGCTGTGAGGCGCTGTGAGGCGCTGTGAGGCG 2580  
Db 4518 GCGCGGTGACGAGCGCGGACGCTGTGAGGCGCTGTGAGGCGCTGTGAGGCGCTGTGAGGCG 4577  
Qy 2581 CTGTGCTTCTGTGGCGCGCTTTCCTTCAGGGGGGCGCGGAGTGTGCGCTGCGGACGTAAC 2640  
Db 4578 CTGTGCTTCTGTGGCGCGCGCTTTCCTTCAGGGGGGCGCGGAGTGTGCGCTGCGGACGTAAC 4637  
Qy 2641 CTTTGGGAGCGCGGAGCTGTGATGACACGAAAGCGAGCGCGCGCGCGCGCGAGC 2700  
Db 4638 CTTTGGGAGCGCGGAGCTGTGATGACACGAAAGCGAGCGCGCGCGCGCGAGC 4697  
Qy 2701 GCGCGGTGCTCGGAGCGGCTGTGACGAGAGTGTGAGAGGAGGAGGCGCGGTGCGCGCGCG 2760  
Db 4698 GCGCGGTGCTCGGAGCGGCTGTGACGAGAGTGTGAGAGGAGGAGGCGCGGTGCGCGCGCG 4757  
Qy 2761 GACGCGGAGCGCTGTGCTGTGACATGCGCGCGCGGAGCGAGCGCGCGGAGAGGTC 2820

Db 4758 GACGCGGAGCGCTGTGCTGTGACCATGCGCGCGCGCGGAGCGGACCGCGGAGAGGTC 4817  
Qy 2821 GAGCGCGCGCGGAGCGCTGTGCTGTGAGTGTGAGTGTGAGCGAGCGGTGCTGTGATCAC 2880  
Db 4818 GAGCGCGCGCGGAGCGCTGTGCTGTGAGTGTGAGTGTGAGCGAGCGGTGCTGTGATCAC 4877  
Qy 2881 CTGTGTCTTGTGAGTGTGAGGAGCGCGCGCGCTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2940  
Db 4878 CTGTGTCTTGTGAGTGTGAGGAGCGCGCGCGCGCTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4937  
Qy 2941 GAGCGCGCGGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3000  
Db 4938 GAGCGCGCGGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4997  
Qy 3001 CTGCGGAGGAGGAGCGCGCGCTGTGCTGTGAGGAGGAGTGTGAGGAGGAGTGTGAGGAGGAGTGTG 3060  
Db 4998 CTGCGGAGGAGGAGCGCGCGCTGTGCTGTGAGGAGGAGTGTGAGGAGGAGTGTGAGGAGGAGTGTG 5057  
Qy 3061 GCGGTGAGCGAGGAGCGGTGAGCGGCTGTGTGAGGAGCGCAACCGGTGATGCGCTTTCGCGG 3120  
Db 5058 GCGGTGAGCGAGGAGCGGTGAGCGGCTGTGTGAGGAGCGCAACCGGTGATGCGCTTTCGCGG 5117  
Qy 3121 GAGCGCTTGTGTACCAAGTGTGAGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3180  
Db 5118 GAGCGCTTGTGTACCAAGTGTGAGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5177  
Qy 3181 CTCTCGGCGATGAGGAGCGCGCGCATGCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3240  
Db 5178 CTCTCGGCGATGAGGAGCGCGCGCATGCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5237  
Qy 3241 GACGAAATAGCGCGCTTTCAGCGCGGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 3300  
Db 5238 GACGAAATAGCGCGCTTTCAGCGCGGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 5297  
Qy 3301 GTGCGTGTGCGCGCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3360  
Db 5298 GTGCGTGTGCGCGCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5357  
Qy 3361 GGCAGCGCGGAGGAGCGCGCTTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3420  
Db 5358 GGCAGCGCGGAGGAGCGCGCTTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5417  
Qy 3421 GCTGTGAGCGCGGTGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3480  
Db 5418 GCTGTGAGCGCGGTGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5477  
Qy 3481 GTGCTCAATCTGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3540  
Db 5478 GTGCTCAATCTGTGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5537  
Qy 3541 GCGCGGTGTGTGAGCTGTGAGGAGCGGAGCTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTG 3600  
Db 5538 GCGCGGTGTGTGAGCTGTGAGGAGCGGAGCTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTG 5597  
Qy 3601 CCGTTCCTGTGAGCAATCTTCTCTCTGTGTGTGATCTCCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3660  
Db 5598 CCGTTCCTGTGAGCAATCTTCTCTCTGTGTGTGATCTCCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5657  
Qy 3661 CCGGCGCGGAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3720  
Db 5658 CCGGCGCGGAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5717  
Qy 3721 ACCGCTCCCGCATGTGAGGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3780  
Db 5718 ACCGCTCCCGCATGTGAGGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5777  
Qy 3781 GCGGAGGCGAGCAATCTTGTGAGAGTGTGATCTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3840  
Db 5778 GCGGAGGCGAGCAATCTTGTGAGAGTGTGATCTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5837  
Qy 3841 GGTATTTCCAACTTCAAGCGAGGCGCGCGCTGTGACCGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTG 3900  
Db 5838 GGTATTTCCAACTTCAAGCGAGGCGCGCGCTGTGACCGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTG 5897



Db 43974 GGGGGTCTTCGTCGCTTTACAGAGGACTACTCTCTGTTGTC-----TCTCGGA 44027  
Qy 441 GTCCGAGAGATGACGCTCATGAGCGGGCTGGGGAGCATGCGCAGCGTCGAGCGGGCG 500  
Db 44028 TCTCAAGCCATCAACGTCCTTACCGGACAGGACGCGCCAAACAGCGGGCTGGAGCG 44087  
Qy 501 AATCTGTATGACCTCTGCGGCTGCGAGGACGCTGTGTGCGGGTGAATACGGCTTATCTGC 560  
Db 44088 ATTGCTCATCTGCTGAGCTGCAAGGGCCAGCTTTGTGGTGAACAAGGCTGTCTCTC 44147  
Qy 561 CTGCTGTGTCCTTTCACTGTGCTGTCAAGCTTGGCTTCGGGGATGCTCCAGCGC 620  
Db 44148 CTCCTTGTCCCGTTCACCTGCGGTGTCAAGGCTTGTGAATGCTGTATCAGATGCG 44207  
Qy 621 CTTGCTGTGTTGGGATGCTGTATGTTGTGCGCCAGACCCCTGTGTGCTCTGAGAC 680  
Db 44208 GGTGTGGGCGGCTGACGTTGTGTCTACGCTACCTGACCTCTGTATGCGGTGAAAT 44267  
Qy 681 CCGGCGCTGCGCAGGACGCTGCTGCAAGGCAATTTTCGCGAGGCGCGATGGGTTGCG 740  
Db 44268 GCACGTATGCGCGCGAGCGCGCGCTGCAAGCTTGCAGACGCGCGCGATGGATGCG 44327  
Qy 741 ACGAGCGAAGGTTGCGCGCTGTGTCTTCAAGCGCTCACTGAGAGCCGCGCGAGCG 800  
Db 44328 CATGGGTGAGGCTGTGCGGTGTGTATCTCAAGCGCTTGTGAGATGCGCAGCGCTGAG 44387  
Qy 801 CGATCGATATTGGCGGTGATTCGAGATCGCGCATCAATCAAGCGGTGGAGAGCGCG 860  
Db 44388 TGAATCGATCTGTGGGTGATCAAGCTTCGCGCATCAATCAAGCGCGCGCATATGG 44447  
Qy 861 TCTGACCGTGCAGACGAGAGCTCCCAAGAAATGCTGCTGAACGCGGCTTGCGGAGCG 920  
Db 44448 GCTAAGACGCGCGAGTGGCGGCTTCCAGAGGTGTGTGGGTGAGGCGCTGAGCGCGCG 44507  
Qy 921 AGGCTGCGCGGCTTCTGTTGGTGTATGTGAGGACACAGCGACGCGGACGAGCTTGG 980  
Db 44508 TGGCATCTGCGCTGCGAGGTGACGTACGTGAAGCCACAGGTTCGAGGAGACACGCTGCG 44567  
Qy 981 TGACCCCATGGAATCCAAAGCTGGAATGCGGTATACGGGCTCGGCGAGATGTCGAC 1040  
Db 44568 CGATCCGATGAGGTGAGGCGCTTCGCGAGGTATCTGGGACGACCACTTCGCGGCG 44627  
Qy 1041 GCGCTGCTGATCGGCTGCTGAGAACCAACCTTGGCCATCTTGAATGCTGCGGAGAT 1100  
Db 44628 TCCGTTGCTCTGGGTTCCGTAAGGCCAATCTGGGACATACCGCGCGGCGGCGGCGT 44687  
Qy 1101 CACTGGGCTGCTGAAGTGTCTTGTCTCTTACGACGCGGAGATTTCTTGGCACTTCA 1160  
Db 44688 CGCGGGGCTGATCAAAAGTGTGTCTCTCATGACAGAACGAGCGATCTCTCAACTGCA 44747  
Qy 1161 CGCGAGGCGCTGAACCCCGGATCTCATGGGGTGAATTTGGCGGACCGTCAAGCGCGC 1220  
Db 44748 CTACCGGAGCTCAACTCAACATCTCTTGAAGAACCTTCCTGTGATTTCCGAGGA 44807  
Qy 1221 CCGGACACCTGTGCGCGGACTGGAATACGCGCGACGCGGGGGGTGAGCTGTTCGAGAT 1280  
Db 44808 GCTGCATCTCTGGCTGCGGGGAAAGGGGCCCGGTATCCGAGGATCAATGCTTCGTGG 44867  
Qy 1281 GAGCGGACCAACGCGACGTTGTGTGTGGAAGAGGCGCGCGGACGTCACACGCGC 1340  
Db 44868 GTCAAGGACCAACGCGACGTTGTGTGTGGAAGAGGCGCTCGCGCCCAAGAGCTCGCGC 44927  
Qy 1341 GCGCGCGAGGACGCGAGAGCTGTGTGTGTGCGGACGCGGTCGAGCGCTGAGCGCG 1400  
Db 44928 GCGCGTTGAGGCTGTGCGGACCTCTCTACCTGTGCGCTGAGACGCGGCGCTGCG 44987  
Qy 1401 TGCAAGGCGCGCGGCTGTGCGGACCATCTGAGAGCTTACCTTGCAGTGTCTGAGCGA 1460  
Db 44988 CGAGTTGCGCGAGGTGTGAAGACATATGTCTCGCGGACGCGGATCTGCCCATGCGCGA 45047  
Qy 1461 TGTGCGTTCACTGTGCGAGAGCGCGGACGAGTGAAGCACTCGGCTCGCGTGGCGCG 1520

Db 45048 TCTCTGCTACTCGGGAACACGCGGGTGTGGAATGAGACCATCTGCGGTCGCCGC 45107  
Qy 1521 GACGTGAGGAGGGGCTGCGGCGAGCCTTGAAGCTTGCAGCGGAGGAGACAGATGCGC 1580  
Db 45108 GGGGACGACGGCCCAATTTGAAGCGAGCTTCTCCGCTTTCTTGAAGGAGACCGCGAAC 45167  
Qy 1581 CGGTGCGGTGCGAGTATCGCCGATCTCTCAACGCGGCAAG--CTGCGCTTCTTTCAC 1637  
Db 45168 AGGACGTCGTAACGAGACAGTCAAGGACGCGGTGGCGCGGAGATCGGCTTCTGTAC 45227  
Qy 1638 CGGACAGGGGCGGAGACGCTGAGGATGGGCGGTGGGCTGTACATGATGATCGGCTT 1697  
Db 45228 GGGGACAGGCTCGAGATACCGCGGATGGGCGGCACTTACAGACAGCCCACTT 45287  
Qy 1698 CCGGAGGCGGTTGACCTGTGACGCTGTGCAACAGGAGCTGACGCGGCGCTCG 1757  
Db 45288 CCGGCGTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 45347  
Qy 1758 CGAGGTATGTGGCGCAACCGGCGACGCTGACGCGCGCTGTGACACAGACGCTT 1817  
Db 45348 GTGCTGATCTACCCGAGCG-----GGGCAAGCTTCTCGATGACAGACGCGCTT 45401  
Qy 1818 CACCCAGCGCGGCTTTCACCTTGAATATGCTGCGCGGCTGTGAGCGGCTGTGAGG 1877  
Db 45402 CACGACCGCGGCTTGTGCTGTGAGTGGGCGCTCGGAACTTGGCGCTGTGAGG 45461  
Qy 1878 TGTAGACCGGAGTGTGCTGCGCGGCAATAGCATGATGATGATGATGATGATGATGAT 1937  
Db 45462 GCTTGTACACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 45521  
Qy 1938 GGGGCTTCTCGCTTGAAGACGCGGTCTTCTGTTGAGTGGCGCGGCGCTGATGCA 1997  
Db 45522 TGGGCTGTTCACCTTCAGAAAGGCTTGGCTCATGTCGAGGCGCGCGCTGATGCA 45581  
Qy 1998 GCGCGCTGCGCGCGCGCGGCGATGATGATGATGATGATGATGATGATGATGATGAT 2057  
Db 45582 GTCGTTCCCGCGGCGCGGAGATGCGCGCTTTCACGAGCGGCGCTCTTACCGGCA 45641  
Qy 2058 TGGGCTGCGCGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2117  
Db 45642 GCGCATGCTCTTACGATGACAGGTCTCATGTCGCGCTTCAATGATCTTCCGAGAC 45701  
Qy 2118 GGTGATGCGCGGCGCGGCAACCGTGCATGCGATGCGCGGCGATGCGCGCGCG 2177  
Db 45702 GGTGATCTTGGCGATGAGCAAGCGGTGAGGCGATTCGCGGCGGTTGTCGCGAGG 45761  
Qy 2178 GCGCGAACAAGCGCTTCCAGCTGTGCGATGCTTCACTACCGCTCATAGCGCGGAT 2237  
Db 45762 CGTGAAGTTGCGGCGCTGAGGTCTCCACGCTTCCACTTCCGCTCATGAGACCCAT 45821  
Qy 2238 GCTGAGGCGTTGCGGCGGTGCGGAGTGGTGAAGTACCGGCGCGCTGATGCTCT 2297  
Db 45822 GCTGAGTCTTCAAGCGGCGGCGGAGAGACGCGCTTCACTGCGCGCGGAGATCAGCT 45881  
Qy 2298 GGTGAGCAATCTGAACCGGAAAGCTTCAACAGAGGTGAAGTTCGCGGCGCTATTGGGT 2357  
Db 45882 GATCTCAACCTTACCGG-----TGGGCTCGGAGAGTTCGCGCGGCGGCTTGG 45935  
Qy 2358 GCGCGACGCGGAGAGGTGTGCTTTCGCGGATGAGTGAAGGCGCTGACGCGGCGG 2417  
Db 45936 TCGACATGCGGTGAACCGGCTCGGTTCTCGGATGAGGCGCGCTCAAGGACGTTG 45995  
Qy 2418 TGGGCGACCTTGTGAGGCTGTGCGGATTCGACGCTGCTGCGGCTGTGCTGCTG 2477  
Db 45996 GCTTCTGCTGCGGTGAGATGAGTCCGAAATGCACTGTCTGCGCATGGAACCGGTG 46055  
Qy 2478 CATGCGGAGCGCGCGCGCGCTGTGCAATGCTGCGCGCTGAGCGGTGAGAGCGCG 2537  
Db 46056 TCTGCGGAAAGGCTGCGGCTGTGCTTCTTCCGAAAGGACAAAGAGAGTGGGA 46115  
Qy 2538 GACCTGCTGAGGCGCTGCGGCGGCTGTGAGCGGTGTGAGCTGTGCTTCTGCGCG 2597  
Db 46116 AGTCTCTGCAACGTTGGGTGCGCTCTGTGCGAGGCGTTCCGTTCACTGGGACG 46175

QY 2598 ---CCTCTTCCCTCAGGGGGGGGGGGTGCCTCCGCTCCACGTAACCTTGGCAGGCGCA 2654  
CC 46176 GTTCAACAGCACTACCCCGCTCGGGGGTCTCCCTCCCAATTATCCCTTTACAGACGA 46235  
DB 2655 GCGCTACTGATGACACCAAGACCGACCGGGCGCTGGACCGCGCTGCTCCGGG 2714  
QY 46236 GCGGTTCCGGCTC-----AAGCAGATGATGACAGGGGTGTGGCGCCGACGGCTCAGG 46288  
DB 2715 AGCGGATCAGACGAGTCAAGAGGGGGCGCGGTGCGCGGCGGACCGCGCAGCGC 2774  
QY 46289 TCGTGGGTATCCCTCGTTGGATCGCGCTCGCGCTCGACGAGGACGACACTCTGT 46348  
DB 2775 TCGGCTCCGACCATCCG 2790  
QY 46349 TCGAATCGAAGTTCCG 46364  
DB

RESULT 4  
AAC55785  
ID AAC55785 standard, DNA; 3978 BP.  
XX AAC55785;  
AC 19-JAN-2001 (first entry)  
DT  
XX  
XX Type I polyketide synthase orf8.  
DE  
XX Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
KW chronic obstructive pulmonary disease; respiratory inflammation;  
KW fungicide; pesticide; de.  
XX  
OS Streptomyces lavendulae.  
XX  
XX WO200053737-A2.  
XX  
XX 14-SEP-2000.  
XX PD  
XX 10-MAR-2000; 2000MO-US06394.  
XX PF  
XX 12-MAR-1999; 99US-0266965.  
XX PR  
XX (MINU) UNIV. MINNESOTA.  
XX PA (SHER) SHERMAN D H.  
XX PA (MAOY) MAO Y.  
XX PA (VARO) VAROGLU M.  
XX PA (HEM) HE M.  
XX PA (SHEL) SHELTON P C.  
XX  
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
XX  
XX WPI; 2000-601980/57.  
XX  
XX  
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
PT the molecular basis of mitosome ring system biosynthesis  
XX  
XX  
XX Example 1; Page 251-253; 399pp; English.  
XX  
XX This invention relates to isolated and purified nucleic acid molecules  
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
XX natural products that contain a variety of functional groups, including  
XX amino benzquinone and axiridine ring systems. The S. lavendulae  
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes  
XX spanning 55kb of DNA. The invention includes an expression cassette  
XX comprising a mitomycin biosynthetic gene operably linked to a promoter,  
XX and host cells transformed with the cassette. The nucleotide, and protein  
XX sequences and the transformed host cells of the invention result in  
XX antiasthmatic, anti-inflammatory, cytostatic, immunomodulatory, and  
XX antibiotic activities. The nucleotide sequences are used to elucidate the  
XX molecular basis for the biosynthesis of the mitosome ring system, as well  
XX as to engineer the biosynthesis of novel natural products, e.g.

CC antibiotics, anti-inflammatory agents, anti-cancer agents,  
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic  
CC obstructive pulmonary disease as well as other disease involving  
CC respiratory inflammation, or cholesterol-lowering agents or as crop  
CC protection agents (e.g. fungicides or insecticides) as well as  
CC biopolymers, e.g., in packaging or biomedical applications, or to engineer  
CC PHA monomer synthases. Sequences AAC55782-C55811, AAC55815-C55849 and  
CC AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA  
CC sequences and encoded proteins. Sequences AAC55812-C55814,  
CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the  
CC cloning of the mitomycin biosynthetic genes.  
XX  
SQ Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 other;  
Query Match 19.1%; Score 814.2; DB 21; Length 3978;  
Best Local Similarity 57.6%; Pred. No. 1.1e-121;  
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;  
QY 18 CGAGCGCGACGCCGAATCCGATGCGATGCGTGCAGCGAGTCCGCTGCGCGTGG 77  
DB 90 CGAGGACGCGGCCCGCGAGCCCGCTGCGATGTGGGACATGGCGTCCCTTCCCGGGA 149  
QY 78 CGTGATCGATCGACCGGGTTCTGGACGCTCTCGAGGGCTCGCGGACACCGTGGGGC 137  
DB 150 CGTGCGATCGCGGACGACCTGTGGCAGCTGTGGCGAGGGCGGAGACGCCGTACCGA 209  
QY 138 AGTCCCGCGCGAAC--GCTGGGATGACGACGCGTGTGATCCGACCCCGATGCCCC 194  
DB 210 GTTCCCGCGCGACCGGGGCTGGAGCTGTGAGCGCGCTACGACCCCGAGCGGCGACCC 269  
QY 195 GGGGAGAGCGCCGCTTACCGCGCGCATCTTCTCGTAGCGATGAGCTCTTGCAGCGCTC 254  
DB 270 GGGCAGGACGTACGCGCGCGCCACGCGGCTTCTTCAAGACGCCCGGATTCAGCGCCG 329  
QY 255 CTCTTTCGGCATCTTCGCTCGCGAAGCGCTCGGATGAGACCCCTGACATGCACTTGTCT 314  
DB 330 CTCTTTCGGCATCAGCGCGCGCGAGGCGCTCGCATGAGCCGACGACGCGCATGATCAT 389  
QY 315 GAGGTGTCTGGAGAGGGGCTGAGAGAACCGCGCATGCTCTCATGCGGCTCTGTGGTAC 374  
DB 390 GAGGTCTCTGGAGAGGGGCTTGAAGCAGCGCGGCTTACGCGACACCTTGGGGGCGA 449  
QY 375 GGAACGGGAGTGTTCATCGGATCGGCGCTCGCATATGAGCGCGCTCGCGCAAGC 434  
DB 450 GAGCGTGGCGCTCTTCTGTGCGCTCAACAGACGACATCTGATCAACGTCTGAGCGC 509  
QY 435 GACGCGCTCGCAGAGATGACGCTCATGCGCGGCTGGGAGCATGCCAGCTGAGAC 494  
DB 510 GCGGAGCGTCCGCGAGGCTTC-----ATCGGAGCGGCAACTCCGACGATCTCTC 563  
QY 495 GGGCGCGATCTGTATGCTCTCGGCTGCGAGGGCGGTGTGCGGGTGTGATACGGCCTA 554  
DB 564 CGGCGCGCTCGCTACACCTTGTGAGAGGGCGCGCGCTGTCTGTGACACCGCCTG 623  
QY 555 TTCGCTCTGCTGAGGCGCTTCACTGCTGCTGTCAGAGCTTGCCTCGCGGAAATGCTC 614  
DB 624 CTCCTCTCTGCTGTGCGGCTGACCTGCGCGGAGTCCCTGCGGAGGGGAAATGCTC 683  
QY 615 CACGCGCTGTGCTGAGGGGATGCTGTGATGTTGTGCGCGAGACCTCTGTGTGCTCTC 674  
DB 664 CTTGCGCTGTGGGGGGCGCGCGACGATGATGCGCCGCGACCGCTTATCATGATTTAG 743  
QY 675 GAAAGCCCGGCGCTGTGCGAAGAGCGTCTGTGCAAGCATTTTGGGGGAGCGCGATGG 734  
DB 744 CCGCAGGCGGGGCTCGGGCGCGCGAGCGCGCTGCAAGTCTTCTGCGGACCGCGCAGG 803  
QY 725 GTTGCAGAGGCGGAAGGGTGGCGCTGTGCTCTCAAGCGCTCACTGAGAGCCCGCGC 794  
DB 804 CACCACTGTGTCCAGGCGCGCGCGCTGTCTGTGTGCGCGGCTCTGAGAGCCCGCGC 863  
QY 795 GAGCGCGATCGGATATTGCGCGGTATTCAGAGATCCGATCAATCAAGCGGTGCGAG 854  
DB 864 CTTGGGCTAACCCCGTGCACGCGGCTCATCCGGGGGAGCGCCCGTCAACGAGAGCGCGG 923

QY 855 CAGCGGTCTGACCGTGCAGGAGCTCCCAAGAAATCGTGTGAAACGAGCCCTTGGC 914  
 DB 924 CGCGGCGCTGACCGGGCCCAACGAGACGAGCGGAGTGAATCCGACAGGCACTGGC 983  
 QY 915 GAGACGAGGCTGCGCGCGCTCTTCGTGGTATATGTGAGGACACGCGACGAGGACGAC 974  
 DB 984 CAAAGCAGCGGTGACGCGCGACAGCGTGCAGCGCTTCAGGACACCGCACCGGACCGCC 1043  
 QY 975 GCTGTGTGATCCCATCGAAATCGAAGCTGTGATGCGGTATACGCGCTTCGAGGAGAT 1034  
 DB 1044 GCTGGGCGAATCCGATCGAGGCGCCAGGCGCTCTCGCACCTACGAGGCGGCGCGGCGA 1103  
 QY 1035 CGCGACCGCGCTGTGATCGGAGTCGAGTAAAGACAACTTGGCCATCTCTGATATGCTC 1094  
 DB 1104 GGGGAGGCGCGCTGTGGCTGGGTCTGCTGATGAGTGAACCTTGGGCGACACCGAGTCCGGGC 1163  
 QY 1095 GGGGATCACTGGGCTGTGAAAGTGTCTTGTCTTTCAGACGAGGACAGATTCTTGGCA 1154  
 DB 1164 CGGCGCGGCGGCGCTCATCAAGATGTATGCGCATGCGGACGAGGACGCTGCCCGGAC 1223  
 QY 1155 CCTCACCGCGAGGCGGTGAACCCCGGATCTCATAGGAGTATCTTCGGCTGACCGTAC 1214  
 DB 1224 GCTGACCTCAACGAGCGCCACCGCGCGTGAATGTCGCGCGGTGACGTACCGCTGCT 1283  
 QY 1215 GCGCGCCCGGACACCGGTGCGGACCTGGAATACGCGCGGAGCGGAGGTGAGCTGCT 1274  
 DB 1284 GACCGAGGCCAGAGACTGGCCGAGACACCGGACGCGCGCGTGGCGCGCTCTCTT 1343  
 QY 1275 CGGCAATGAGCGGAGCAACGCGCACGTGTGTGAAAGGCGCGCGCGACGCTGAC 1334  
 DB 1344 CGGCGTCAAGCGGACCAACGCGCATGTATCTTGAAGGCGCGCGCGCGAGAGGCAAC 1403  
 QY 1335 ACCGCGG-----GCGCGGAGGACCGGACAGGCTGTGTGTCTC 1376  
 DB 1404 GAGCGCGCGCTGCGGACGTCTCTCGACGCGCGGCGCGCTGCGGTGTGTCTC 1463  
 QY 1377 GGAAGGACCGCGTCAAGCTTGAATGACAGGCGGCGCGCTGCGGACCACTTGAAGAC 1436  
 DB 1464 CGGCGCGAGCGAGCGCGCTGTCCGAGCGGCGGCGGCGCGCTGCGGCGCGCTGCGG 1523  
 QY 1437 CTACCTTTCGAGTGTGTGGGCGATGTGGCGCTTCACTGTGGACGACGCGGACGCT 1496  
 DB 1524 GCGCGCGGACCTGCGACCGGCGGAGTGGACACCGCGCTGGCACACGCGGCGCGCT 1583  
 QY 1497 GGAAGACCGGCTCGCGGTGGCGGACGTCAGAGGAGGCGGTGGCGGACCGCTGGAGC 1556  
 DB 1584 CGACCAACGCGCGCGCTGTGTGGCGGAGCGTGAAGAACTGTGCGCGGCTTGGCGG 1643  
 QY 1557 TGGCGCGGAGGAGACAGAGTGCAGCGCGGTGGCGGAGTATCGCGGATCTCTCAACGCG 1616  
 DB 1644 CTGTGCGACCGGACCGCGCGCGCGCGCTGTGTACCGGCGGACCGCGCGCTGCGGCG 1703  
 QY 1617 CAAGCTGCGCTTCTTCTTCAACGAGACAGGAGCGGACGCTGGGACATGGGCGTGGGCT 1676  
 DB 1704 CAAGGCGCGCTTCTTCTTCAACGAGACAGGAGCGGACGCGGACATGGGCGGAGACT 1763  
 QY 1677 GTACGATGATGTGTGGCGCTTCCGAGAGCGTTCGACTGTGCGTGAAGCTTTAAACA 1736  
 DB 1764 GAGCGCTCAACGACCGGTGTGTGCGAGCGCTTGAACGAGTGTGCGGCGGCTGAGACG 1823  
 QY 1737 GAGAGCTGACCGGCGCTCCGAGAGTGTGGCGGAAACGCGGCGGACGCTGAGCGCG 1796  
 DB 1824 GCACTCTGACCGGCGCGCTGCGGAGGTGTGTTCGCGCGGACGCGACCGCGGAGCGCG 1883  
 QY 1797 GCTGTCTGACGACGAGCTTTCACACGCGCGCGCTGTTCACCTTGAATATGCGCTGC 1856  
 DB 1884 CTGTCTGACACGAGCGGCTTACACGAGCGCGCTGTTCGCGCGTGAAGTGTGCGCTGCT 1943  
 QY 1857 GCGCGTGTGGCGGTGTGGGTGTAGAGCGGAGTGTGTGCGCGGCTATGAGATGTGTGA 1916  
 DB 1944 GCGGCTGTGAGAGACTGGGCGCTTGGCGCGCGGACGAGTGTGTGCGGCGCACTGCTGCGG 2003

QY 1917 GCTGTGTGTGCTGCGCGGTGGCGGCGGTGTCTCGCTTGAAGACGCGGATTTCTGTGGC 1976  
 DB 2004 ACTGACCGCGCTTACCGCGCGCGGAGTGTGGTCTGCTGCGCGGACCTGCGCGCTGTGGC 2063  
 QY 1977 TGGCGCGGCGCTTATGACAGGCGCTTGGCGGCGCGCGGCGGCGGCGGCGGCGGCG 2036  
 DB 2064 CGCGCGGCGCGGTGACCGAGGACATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 2123  
 QY 2037 GCGCGGCGGCGATGTGTGCTGTGGGTGGCGCGCGGACGACGAGTGTGTGATGCGCGC 2096  
 DB 2124 GACCGAGGACGAGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2183  
 QY 2097 GGTCAACGCTTCGAGACGAGTGTATCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 2156  
 DB 2184 CGTCAACGAGACGGAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2243  
 QY 2157 GCGCGGATGAGCG 2216  
 DB 2244 GCGGAGTGGCG 2303  
 QY 2217 CTGACCGCTCATGCGCGCGCGGATGTGAGGCGGTTGCGGCGGTGTGGCGGAGTGTGAGCTA 2276  
 DB 2304 CTGCGCGCACTTGAACCGCATGACCGAGCGGTGTGCGGAGTGTGCGAGGAGGTGTCTTA 2363  
 QY 2277 CGGCGCGCGGTGATGTGTCTGTGTGACCAATCTGAGCGGAGAGGCTTGCACAGAGAGT 2336  
 DB 2364 CAGCGCGCGGTCTCTCCCGGAGTGTCTCAAGCTCACGCGGCGCGCGCGCGCGCGCGCG 2423  
 QY 2337 GAGCTCGCGCGGCTATTGGGTGTGCGCACGCGCGAGAGTGTGTGTGTGTGTGTGTGT 2396  
 DB 2424 CGGAGGCGCGGACACTGGGTGTGCGCACGCTCGGAGAGCGGTGTGTGTGTGTGTGTGT 2483  
 QY 2397 GAAGGCGGTGACGCGCGCGGTGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2456  
 DB 2484 CGGCGCGGTGTGCGGCG 2543  
 QY 2457 GCTGCGCGGTGTGCTGCTGCTGATGCGGACGCGCGCGCGCGCGCGCGCGCGCGCG 2513  
 DB 2544 GAGCGCGCGCGGACGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2603  
 QY 2514 GCGCGCTGGGCGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2573  
 DB 2604 GCGCGCGCGGCGCGGCGCGGACCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2663  
 QY 2574 CGGCGCGGT 2633  
 DB 2664 CGGCGCTTGTGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2723  
 QY 2634 CACGTACCTTGGAGCGGCGGAGCGCTACTGTGATGACACGAAAGCGG 2680  
 DB 2724 CACGTACGCTTTCAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2770  
 RESULT 5  
 AAC55840  
 ID AAC55840 standard; DNA; 12249 BP.  
 XX  
 AC AAC55840;  
 XX  
 DT 19-JAN-2001 (first entry)  
 XX  
 DE Complete Mitomycin ORF 1-9 nucleotide sequence.  
 XX  
 KM Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
 KM anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
 KM chronic obstructive pulmonary disease; respiratory inflammation;  
 KM fungicide; pesticide; ds.  
 XX  
 OS Streptomyces lavendulae.  
 XX  
 FN WQ200053737-A2.  
 XX  
 PD 14-SEP-2000.



XX 10-MAR-2000; 2000MO-US06394.  
 PF Best Local Similarity 57.6%; Score 814.2; DB 21; Length 12249;  
 PR 12-MAR-1999; 99US-0266965; Pred. No. 1e-121; Mismatches 1108; Indels 30; Gaps 4;  
 XX (MINU) UNIV MINNESOTA.  
 PA (SHER) SHERMAN D H.  
 PA (MAOY) MAO Y.  
 PA (VARO) VAROGLU M.  
 PA (HEM) HE M.  
 PA (SHEL) SHELTON P C.  
 XX  
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
 DR WPI, 2000-601980/57.  
 XX  
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
 PT the molecular basis of mitosome ring system biosynthesis  
 XX  
 PS Claim 26; Figure 22; 399bp; English.  
 CC This invention relates to isolated and purified nucleic acid molecules  
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
 CC natural products that contain a variety of functional groups, including  
 CC amino benzoguinone and axiridine ring systems. The S. lavendulae  
 CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes  
 CC spanning 55kb of DNA. The invention includes an expression cassette  
 CC comprising a mitomycin biosynthetic gene operably linked to a promoter,  
 CC and host cells transformed with the cassette. The nucleotide, and protein  
 CC sequences and the transformed host cells of the invention result in  
 CC antitumor, antiinflammatory, cytostatic, immunomodulatory, and  
 CC antibiotic activities. The nucleotide sequences are used to elucidate the  
 CC molecular basis for the biosynthesis of the mitosome ring system, as well  
 CC as to engineer the biosynthesis of novel natural products, e.g.  
 CC antibiotics, anti-inflammatory agents, anti-cancer agents,  
 CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic  
 CC obstructive pulmonary disease as well as other disease involving  
 CC respiratory inflammation, or cholesterol-lowering agents or as crop  
 CC protection agents (e.g. fungicides or insecticides) as well as  
 CC biopolymers, e.g., in packaging or biomedical applications, or to engineer  
 CC PHA monomer syntheses. Sequences AAC55782-C55881, AAC5815-C55849 and  
 CC AAB32485-832542 represent mitomycin biosynthetic gene cluster DNA  
 CC sequences and encoded proteins. Sequences AAC55812-C55814,  
 CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the  
 CC cloning of the mitomycin biosynthetic genes.  
 CC  
 CC Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 other;  
 SQ

QY 315 GGAGGTGCTGGAGAGCGCTGAGAACGCCGCGATCGCTCCATCGCGCTCGTGTAC 374  
 DB 1377 GGAGGTCTCTGGAGAGCGCTTTCAGACAGCGCGCTTCAGCGGACCTTCGCGGAGCA 1436  
 QY 375 GGAAACGGAGAGTGTTCATCGGAGATCGCGCGCTCCGAAATATAGAGCGCGCTCCGCAAGC 434  
 DB 1437 GGACGTGCGCGCTTCGTTCGTGGCTCCACAGCAAGACTATCACTGATCAACGTGTGACGCGC 1496  
 QY 435 GACGCGCTCCGACAGATGACAGCTCATGCGCGCGCTGGGAGAGATGCCACGTGTGAGC 494  
 DB 1497 GCGGACGTCTCCGAGGCGCTTC-----ATCGGACCGGCAACTCCGCGACATCTCTTC 1550  
 QY 495 GGGCGGAATCTGTATGCGCTCGGCGTGGAGAGCGCGCTGTGCGGAGATATGACGCGCTA 554  
 DB 1551 CGCGCGCTCGCTTACACTTCCTGAGAGCGCGCGCTTCGCTGCGACCGCGCTG 1610  
 QY 555 TTCGCTCTGCTGAGCGCGCTTCATCTGAGCTGTGACAGCTTTCGCTCCGCGGATGCTC 614  
 DB 1611 CTCCTCTCTGCTGCTGCGCTGACCTGCGCGGAGTCCCTGCGGAGAGGAGTGTCTC 1670  
 QY 615 CACGCGCTGCTGCTGAGGTATGCTGTATGTTGCTCGGACACCTTCGTGTGCTCTC 674  
 DB 1671 CCGCGCGCTGCGCGCGCGCGCGAGCGGTATGCGCGCGCGCTTCATGATGATG 1730  
 QY 675 GAAGACCGCGCGCGCTGCGCGAGGACGCTGATGCGCGCGCGCTTCATGATGATG 734  
 DB 1731 CCGCGAGCGCGCGCTGCGCGCGCGCGCGCGCGCTGCGCGCGCGCTTCATGATGATG 1790  
 QY 735 GTTTCGAGAGCGAGAGGTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 794  
 DB 1791 CACGACCTGCTGCTGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1850  
 QY 795 GACGCGCGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 854  
 DB 1851 CCGTGGCTTACCG 1910  
 QY 855 CAGCGCTGACCG 914  
 DB 1911 CGCGCGCTTACCG 1970  
 QY 915 GACGCGAGCTGCGCGCGCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 974  
 DB 1971 CAGCGCAGCGCTGACG 2030  
 QY 975 GCTTGTGATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1034  
 DB 2031 GCTGCGCAGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2090  
 QY 1035 CGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094  
 DB 2091 GCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2150  
 QY 1095 GGGGATCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1154  
 DB 2151 CGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2210  
 QY 1155 CTTCCACGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1214  
 DB 2211 GCTGCACTTCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2270  
 QY 1215 GCGCGCGCGCGACCGTGGCGCGAGCTGGAATACCGCGCGAGCGCGGAGTGTGCTGCT 1274  
 DB 2271 GACCGAGCGCGAGACTGCGCGCGAGACCGGACCGCGCGCGCGCGCGCGCGCGCGCT 2330  
 QY 1275 CGGATGAGCGGAGCAACCGCGCGAGTGTGCTGGAAGAGCGCGCGCGCGCGCGCGCG 1334  
 DB 2331 CGCGCTGACGCGCACCAACCGCGCGAGTGTGCTGGAAGAGCGCGCGCGCGCGCGCGCG 2390  
 QY 1335 ACCGCG-----GCGCGAGAGCGCGCGAGCGCGAGCGCTGCTGCTGCTGCTGCT 1376  
 DB 2391 GAGCGCGCGCGTGCAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2450  
 QY 1377 GGCAGAGACCGCGCTGACCTGTGATGCAAGCGCGCGCGCTGCGCGACATCTGAGAGC 1436

Db 2451 CGGCGCAGAGGCGGCGCTCCGAGCCGACGCGCGCTGCGCGCCCACTTGACCGC 2510  
 GY 1437 CTACCTTCGAGTGTCTGGGGGATGGCGCTTCACTGTGGGAGACAGCGAGCGCAT 1496  
 Db 2511 GCGCCCGCACTTGACCGCGCGAGTGGCCACCGCGCTGGCACACAGCGCGCGCTT 2570  
 GY 1497 GAGACACCGGCTCGCGTGGCGGACGTCGAGGAGGAGGCGCTGCGGAGCGCTTGAGCG 1556  
 Db 2571 CGACACCGCGCGCGCTGCTGTCGCGCGGAGCCGTGAGGAACTGCTGCGCGCGCTGCGCG 2630  
 GY 1557 TCGCGCGAGGACAGACGTGCGCGCGTGGCGGTGCGCGCATATCGCGATTCTTCA 1616  
 Db 2631 CTGCGCGACCGGAAACCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2690  
 GY 1617 CAAGTCGCGCTTCTCTTCAACCGGACAGGCGCGCGGACAGCGCTGCGCGCGCTGCG 1676  
 Db 2691 CAAGCGCGCTTCTCTTCAACCGGACAGGCGCGCGGACAGCGCTGCGCGCGCTGCG 2750  
 GY 1677 GTACGATGATGATCGCGCTTCCGCGAGGCGCTTCACTGCTGCTGCTGCTGCTGCT 1736  
 Db 2751 GCGCGCTCAGACGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2810  
 GY 1737 GAGCTCAGACCGCGCTTCCGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796  
 Db 2811 GCACCTCAGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2870  
 GY 1797 GCTGTGACGACGACGCTTCAACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1856  
 Db 2871 CCTGTGACGACGACGCTTCAACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2930  
 GY 1857 GCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1916  
 Db 2931 GCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2990  
 GY 1917 GCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1976  
 Db 2991 ACTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3050  
 GY 1977 TGGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2036  
 Db 3051 GCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3110  
 GY 2037 GCGCGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2096  
 Db 3111 GACCGAGGACGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3170  
 GY 2097 GATCAACGCTCCGACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2156  
 Db 3171 CGTCAACGAGACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3230  
 GY 2157 GCGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2216  
 Db 3231 GCGCGAGTGGCG 3290  
 GY 2217 CTCAACCGCTCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2276  
 Db 3291 CTCCGCGCACTGAGCGCATGACCGAGCGCTTCCGCGAGTCCGACAGAGGAGTCT 3350  
 GY 2277 CCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2336  
 Db 3351 CAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3410  
 GY 2337 GAGCTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2396  
 Db 3411 CCGAGCGCGCGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3470  
 GY 2397 GAGCGCGCTGACGCGCGCGCTGCGGACCTTCTGCTGAGTGGTCCGAATCGACGCT 2456  
 Db 3471 CCGCGCGCTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3530  
 GY 2457 GCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2513

Db 3531 GACGCGCGCGACGCGCGAGTCTGCGCGACCGCGCGCGCGCGCGCGCGCTGCTG 3590  
 GY 2514 GCGCGCTGAGGCTGACGAGCGCGCGACGCTGCTGAGGCGCTCGAGCGCGCTGCG 2573  
 Db 3591 GCG 3650  
 GY 2574 CCGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2633  
 Db 3651 CCGCGCTGCG 3710  
 GY 2634 CACGTACCTTGGGACGCGCGAGCGCTGCTGAGTGAACAGGAAACCG 2680  
 Db 3711 CACGTACCGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3757

RESULT 6  
 AAC5857  
 ID AAC5857 standard; DNA, 18331 BP.  
 AC AAC5857;  
 DT 19-JAN-2001 (first entry)  
 DE Complete nucleotide sequence of the mitomycin gene cluster.  
 KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;  
 KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW fungicide; pesticide; ds.  
 OS Streptomyces lavendulae.  
 PN WO20053737-A2.  
 PD 14-SEP-2000.  
 PF 10-MAR-2000; 2000WO-US06394.  
 PR 12-MAR-1999; 99US-0266965.  
 PA (MINTU) UNIV MINNESOTA.  
 PA (SHER) SHERMAN D H.  
 PA (MAOY) MAO Y.  
 PA (VARO) VAROGLU M.  
 PA (HEM) HE M.  
 PA (SHEL) SHELTON P C.  
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;  
 DR WPI, 2000-601980/57.  
 XX  
 XX  
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene  
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating  
 PT the molecular basis of mitosome ring system biosynthesis  
 XX  
 PS Disclosure; Figure 21; 399pp; English.  
 CC This invention relates to isolated and purified nucleic acid molecules  
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of  
 CC natural products that contain a variety of functional groups, including  
 CC amino benzoguinone and axiridine ring systems. The S. lavendulae  
 CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes  
 CC spanning 55kb of DNA. The invention includes an expression cassette  
 CC comprising a mitomycin biosynthetic gene operably linked to a promoter,  
 CC and host cells transformed with the cassette. The nucleotide, and protein  
 CC sequences and the transformed host cells of the invention result in  
 CC antiasthmatic, anti-inflammatory, cyostatic, immunomodulatory, and  
 CC antibiotic activities. The nucleotide sequences are used to elucidate the  
 CC molecular basis for the biosynthesis of the mitosome ring system, as well  
 CC as to engineer the biosynthesis of novel natural products, e.g.  
 CC antibiotics, anti-inflammatory agents, anti-cancer agents,  
 CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic  
 CC obstructive pulmonary disease as well as other disease involving

CC respiratory inflammation, or cholesterol-lowering agents or as crop  
CC protection agents (e.g. fungicides or insecticides) as well as  
CC biopolymers, e.g., in packaging or biomedical applications, or to engineer  
CC PHA monomer synthases. Sequences AAC55782-C55881, AAC5815-C55849 and  
CC AAB32485-832542 represent mitomycin biosynthetic gene cluster DNA  
CC sequences and encoded proteins. Sequences AAC55812-C55814,  
CC AAC5850-C55856 and AAC5862-C55869 represent PCR primers used in the  
CC cloning of the mitomycin biosynthetic genes.

XX Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 other;

Query Match 19.1%; Score 814.2; DB 21; Length 18331;  
Best Local Similarity 57.6%; Pred. No. 1e-121;  
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

```
QY 18 CGAGCGCGCAGCCGAAATCCGATTCGATCGGAGCGAGTTCGCTGCCCCGGTGG 77
DB 1077 CGAGGACCGGCGCCGCGAGCCCGTCGATCGTGGCATGGCGTCCGCTCCCGGGGA 1136
QY 78 CGTGATCGATTCGAGGGGTTCTGACGCTCTCGAGGGCTCGCGACACCTCGGGCG 137
DB 1137 CGTGCGATCGCGGAGGACCTGTGGCGAGTGTGCGGAGGGCCGGGACGCGTCACCGA 1196
QY 138 AGTCCCGCCGGAAC---GCTGGGATCGACAGCGGTGTTGATCCCGAATCCGATCCCG 194
DB 1197 GTTCCCGCGGACCGGGGGCTGGGACGTCAACCGCTCTACGACCCCGAGCCGGGACCC 1256
QY 195 GGGGAAAGACGCCCGTTACCGCGGCATCTTCTCGAGCGAGTACCTGCTTGAACGCTTC 254
DB 1257 GGGGAGAGAGTACGCGCGCCACGCGGGCTTCTCTCAAGAGACGCGCGGATTCGACGCG 1316
QY 255 CTTTCTTGGCATCTCGCTTCGCGAAGCGCTGCGGATGACCTCGACATCGACTCTTGT 314
DB 1317 CTTCTTCGCGATCAACCGCGCGAGCGCTCCGACATGACCCGACACGCAATGATCAT 1376
QY 315 GGAGGTGCTGAGGAGCGCTGAGAGACCGCGCGATCGCTCCATCGGCGCTGTCGATAC 374
DB 1377 GAGAGTCTCTGGAGGCGCTTGGAGGAGCGGGCTTCGACGCGACCACTTCGGGGCGCA 1436
QY 375 GGAAACGGGAGTTCATCGGAGTGGCCCGCTCGAATATGAGCCGCGCTGCCGCAAC 434
DB 1437 GAGCGTGGGCTTCTGCTGGCTCCAAAGCAAGCATCTGATCAAGTGTCTGACGCG 1496
QY 435 GACGCGCTCCGACAGATGACGCTCATGCGGGGCTGGGAGACGATCCAGAGTCGAGAC 494
DB 1497 GGGGACGTGCGGAGGCTTC-----ATCGGAGACCGGCAATCGCGCAGATCTCTC 1550
QY 495 GGGCGAATCTGTATGCTCTCGGCGTGGAGGCGGCTGTGTCGCGGATATACGAGCTA 554
DB 1551 CGGCGCGCTGCTACACTTGGGCTTCAGAGGGCCCGGCTGTCTGCTGACACCGGCTG 1610
QY 555 TTGCTCTCGTGTGGCGGCTTCATCTGAGCTGTCTGACAGCTTGGCTCGGGGAATGCTC 614
DB 1611 CTCTCTCTGCTGTGTGCGCTGCTCACTTGGCCGCGAGTCTCTGCGGCAAGGGAAGTCTC 1670
QY 615 CACGCGCTCGCTGTGGGGTATCGCTGATGTTGTTCGCGAGACCTCGTGTGTGCTTC 674
DB 1671 CTGTGGCTGGGGGGGCGGCGCAAGTGTATGCGACGCGCTTATGAGTTGAG 1730
QY 675 GAAGACCGCGGCGCTGGCAGGAGCGGTGCGCAAGGATTTTGGCGGAGGCGGATG 734
DB 1731 CGGCGAGGGGGCTTGCCCCCGGACGCGCTGCGAAGTCTTCTGCGGACCGCGCGAG 1790
QY 735 GTTGGACGAGGCGAAAGGTGCGCGCTGTGCTCAAGCGGCTCATGTGAGCCGCGCG 794
DB 1791 CACCACTGTCTCGAGGGCGGCGGCTGTGCTGTGCTGCGCTCTCGGACGCGCGCG 1850
QY 795 GACGCGCATTCGATATTGGCGGTATTTGAGAGATCCGCGATCATCAGAGGTTGCGAG 854
DB 1851 CCTGGGCTTACCCCGTCAAGCGGTATTCGGGGGCAAGCCGCTCAACAGAGAGCGCGAG 1910
QY 855 CAGCGCTGACGCGTCCGAGCGGAGCTCCCAAGAAATCGTGTGAACGGGCGCTGAC 914
DB 1911 CGCGGGCTGACCGCGCCCAACGAGACCGGCGCAACAGCGGGTATTCGGGACGACCTG 970
QY 915 GAGCGCAGGCTGCGCGCGCTTTCGCTGGGTTATGTCAAGGACACCGGACGGGACGAC 974
DB 1971 CAACCCAGCGGCTGACCGCGCGACAGGTGTGACGCGGTTCAGGACACCGCACCGGAC 2030
QY 975 GCTTGTATACCCCATTCGAATTCGAAGCTGTGAATGCGGTATATCGGCTTCGGGCGAGAT 1034
DB 2031 GCTGGGCGACCCGATCGAGGCGCCAGGCGCTCTCTCGCACCTACAGGCGGGCGCGGCGA 2090
QY 1035 CGCCAGCGCGCTGTGATCGGATCGGTGAGAGACCAACTTGGCCATCTGATATGCTC 1094
DB 2091 GGGCAGGCGCGCTGTGGCTGGGCTGTGAAGTCGAACCTGAGGCGCACACCACTGCGGCG 2150
QY 1095 GGGGATCACTGGGCTCTGAAGTGTCTTGTCTTCTTACGACCGGCGAGATTTCTGCGCA 1154
DB 2151 CGGCGCGGCGCGCTCATCAAGATGTATAGCCATGCGGACGCGGACGCTGCCCGGAC 2210
QY 1155 CTTCAAGCGGAGGCGGTGAACCCCGGATCGATGAGGTGATCTTTCGCTGACCGTCA 1214
DB 2211 GCTGCACTTCACGAGGCGCCACCGCGCTGACCTGATGTCGCGGTACGTAAGCTGCT 2270
QY 1215 GCGCGCGCGGACACCGGTGCGGACTGGAATACCGCGGACGCGGCGGGGTGAGCTGTT 1274
DB 2271 GACCGAGGCGCAGACTGTGGCGGACACCGGACAGCGCGCGGTGGCGCTGCTGCTT 2330
QY 1275 CGGCATGACGGGACCAACCGGACGCTGTGTGTAAGAGCGCGCGCGGACGCTGAC 1334
DB 2331 CGGCTGACGCGCACCAACGCGCATGTGATCTTGAAGGCGCGCGCGCGGAGAGGAC 2390
QY 1335 ACCGCG-----GCGCGGAGCGACCGGAGAGCTGTGTGCTGTC 1376
DB 2391 GAGCGCGCGCTGCCGAGCTCTCTGCGACGCGGGGCGCGCTGCTGCTGTGCTTC 2450
QY 1377 GGCAGAGACCGCGTTCAGCTTGAATGCAAGCGGCGCGGCTGCGGACCATCTGAGAC 1436
DB 2451 CGCGCGGACGAGGCGCGCTGCGAGGCGGAGCGGCGGCGCTGCGGCGCTGACCGC 2510
QY 1437 CTACCTTTCGATGTCTGGGCGATGTGCGCTTCACTGTGCGACGACGCGGCGCAT 1496
DB 2511 GCGCGCGACCTGCGACCGGCGGACGTGCGGACCGCGCTGCGACACGAGCGGCGGCTT 2570
QY 1497 GAGACCGCGCTCGGCTGCGGCGGACGTCGAGGAGGCGGCTGCGGAGCGCTGAGAC 1556
DB 2571 GACACCGAGGCGCGCTGCTGCGGCGGACCGGTGAGAACTGCTGCGCGGCTGCGG 2630
QY 1557 TCGCGCGAGGACAGAGCTGCGCGCGGTGCGGTGCGAGTATCGCGATTCCTACGCGG 1616
DB 2631 CTTGCTCACCGGACCGCGCGCGCGCTGCTGACGCGCGGACCCGCGTCCGCGG 2690
QY 1617 CAACTGCGCTTCTTCTTCAACCGGACAGGAGCGGACGCTGCGCATGCGGCGCT 1676
DB 2691 CAAAGCGCGCTTCTTCTTCAACCGGACAGGAGCGGACGCTGCGCATGCGGCGGAACT 2750
QY 1677 GTACATGATATGATCGGCTTCCGAGGCGCTTCACTGTGCGTGAAGGCTGTTCAACA 1736
DB 2751 GCGCGCTACAGACACCGTGTTCGCGAGCGCTTGAACGAGGTCTGCGGCGGCTGAC 2810
QY 1737 GAGCTCGACCGCGCTTCCGCGAGGTATGAGGCGGACCGGCAAGCTGTGACGCGCG 1796
DB 2811 GCACTTCGACCGGCGGCTGCGGAGGTCTGCTGCGCGGACGCGCGCGGAGGCGCG 2870
QY 1797 GCTGCTGACCAAGACAGCTTCAACCGGCGGCGCTGTTCACTTTCGAAATATGCGCTGC 1856
DB 2871 CTTGCTGACCAAGAGGCGCTTCAACCGGCGGCGCTGTTTCGCGGAGGTGCGGCTCT 2930
QY 1857 GCGCGTGTGCGGCTGTGAGGCTGAGAGCGGAGTGTGCTGCGGCGCATATGACGCTGA 1916
DB 2931 GCGGCTGTGAGAGACTGTGGGCTTGGCGCTTGGCGCGGATGTGCGGCGCACTGCTGCG 2990
QY 1917 GCTGTGTGCTGCTGCTGCGGCGGCGGCTTCTGCTTGAAGACGCGGATGCTGCTGCTG 1976
DB 2991 ACTGACCGCGCTTACGCGCGCGGCGGCTGTGCTGCTGCGGACGCGCTGCTGCTG 3050
```

QY	1977	TGCGCGCGGCGCCCTGATGACAGAGCGCTCCGCGCGCGGGCGATAGTGTGCATCGAAGC	203.6
Db	3051	CGCCCGCGCGCGCTGACCCAGGACCTCCCGCGGCGCGCCCATAGTGTGCGTCAGAGC	311.0
QY	2037	GCCGAGGCGCATGTGCTGCTCGGTGAGCCCGACGACGTCGTGATGATGCCGC	209.6
Db	3111	GACCGAGGACAGAGTGCAGCGCCCACTCGCGACGCGCGCCCGCGCTGAGCATGTCGCGC	317.0
QY	2097	GATCAACGCTCCGAGCCAGAGTGTATATCGCGGCGCGCGGACACCGCGTGCATGTGCATGCGC	215.6
Db	3171	GCTCAACGGAACCGAAGCGGTGTCTCTCCGCGACAGAGCCCGCTTCACAGGACTGCGC	323.0
QY	2157	GCGCGCATGAGCCCGCGCGCGGAGCGCAACAGGCGCTCCAGCTTCGACATGCGTTCCA	221.6
Db	3231	GCGCGAGTGGGCGCGCCCGCGCGCGGAGACAGAGAGCTGCGGGTCACGCCACTTCCTCCA	329.0
QY	2217	CTCACCGCTCANNCGCCCGATGCTGTGAGAGCGTTCCGCGGTGTGCGCGCATGGTGAAGCTA	227.6
Db	3291	CTCGCGCCACTGAGCGCCATGACCAGAGCGTTCCGCGAGTGTGACAGAGAGTGTCTTA	335.0
QY	2277	CCGCGCGCGCTGTGATCTGTCTGTCTGACGAACTGAGCGGAGGCTTGCACAGACGAGGT	233.6
Db	3351	CAGCGCGCGTCTCCCTCCCGGTGTGTCTCAGCGTCAACGCGGCGCCCGTTCACCGAGAGCT	341.0
QY	2337	GAGCTCCGCGCGCTATTGGGTGCCGCACACGCGGAGAGGTGTGGCTTCCGCGATGAGGT	239.6
Db	3411	CCGCAAGCGCGGAAACATCGGAGTGTGCGCAAGTCCGGAGACGAGTGTGCTTCCACGACGCGGT	347.0
QY	2397	GAAGCGCTGACAGCGCGCGGTGTGCGGAGCACTTGTGTGACAGGTCCGGAATCGACGCT	245.6
Db	3471	CCGCGCCCTGTGGCCACACCGCGGGCCACCGCGCTTCTGTGAGGTGTGCGGCCCGCGCGTGTCT	353.0
QY	2457	GCTGGCGCTGTGCTGCTGCTGCGATGCGGACGCGCCGCGC---GCGCTGCTCGCATGCTC	251.3
Db	3511	GACGCGCCCGGCACGCGCGATGCTGTCCCGAGCGCGCCGCCGAGACGTTGTGTCCCGTGTCT	359.0
QY	2514	GCGCGCTGAGGAGTGAACGAGCGCGGACAGCTGTGTGAGGCGGTCCGCGCGGAGCTCTGTGGCGGT	257.3
Db	3551	GCGGCGCGCCAGGCGCCGAAACCGAGTCCGTGTGACGCGCTGTGCGCAGGCGCCACACGAT	365.0
QY	2574	CGGTGGCTGTGCTCTGTGGGCGCGGCTTTCCTCCACAGGAGGCGCGCGGATGCGCTGCC	263.3
Db	3651	CGGCTCTCGCGGAGTGGAGACGCGCTCTGCGCCCAAGGCCGGAACGCGCGTGTGACCTGCC	371.0
QY	2634	CACGTACCTTTGGACAGCGCGGACGCTACTGTGATGCAACAGAAAGCGC	268.0
Db	3711	CACGTACCTCTTCAGCGCGGCGCACTTGTGTGTGCGTGGCGGACATGGCGC	375.7

RESULT	7
AAID17185	
ID	AAID17185 standard; DNA, 27541 BP.
XX	
AC	AAID17185;
XX	
DT	29-NOV-2001 (first entry)
DE	
XX	Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
DE	
XX	Polyketide synthase, PKS; macroide; nystatin; PKS gene cluster
KW	antifungal; antibiotic; nys2; ds.
XX	
OS	Streptomyces noursei.
XX	
Key	Location/Qualifiers
FH	complement (454..1191)
FT	/*tag= a
FT	/product= "NysF protein"
FT	/note= "CDS does not include start codon"
FT	complement (1275..3092)
FT	/*tag= b
CDS	/product= "NysG protein"
FT	

	CDS	complement (3070..4824)
FT	/+tag= c	
FT	/note= "CDS does not include start codon"	
FT	/product= "NysH protein"	
FT	5122..6156	
CD	+tag= d	
FT	/product= "NysD3 protein"	
FT	6318..27541	
FT	+tag= e	
FT	/product= "NysI partial protein"	
FT	/note= "CDS does not include stop codon"	
PV	MW020159126-A2.	
PD	16-AUG-2001.	
PX	08-FEB-2001; 2001WO-GB00509.	
PX	08-FEB-2000; 2000GB-0002840.	
PR	10-APR-2000; 2000GB-0008786.	
PR	14-MAR-2000; 2000GB-0009387.	
XX	(UNO-) UNIV NORGES TEKNIISK NATUREVIDENSKAPELIGE.	
PA	(SMTF) SINTEF STIFTELSEN IND TEK FORSK.	
PA	(ALPH-) ALPHARMA AS.	
PA	(SINV-) SINGENT AS.	
PA	(IDIE/) DZIGLEWSKA H.	
PA	(ZOTC/) ZOTCHEV S B.	
PA	(SEKU/) SEKUROVA O N.	
PA	(FUAE/) FJAEVIK E.	
PA	(BRAU/) BRAUTASET T.	
PA	(STRO/) STROM A R.	
PV	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;	
PI	Valla S, Ellingsen TE, Sletta H, Gullfksen O;	
DR	MP1: 2001-557614/62.	
DR	P-PSDBI, AAEE10136, AAEE10139, AAEE10140, AAEE10141, AAEE10142.	
PT	New nystatin polyketide synthase polynucleotides and polypeptides,	
PT	useful as antibiotics and antifungals -	
XX	Claim 2; Page 151-166; 266pp; English.	
CC	The present invention relates to the cloning and sequencing of the gene	
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme	
CC	involved in the biosynthesis of the macrocyclic antibiotic nystatin.	
CC	The nystatin PKS is useful as antifungal antibiotics. The present	
CC	sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.	
SQ	Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;	
Dd	Query Match 18.9%; Score 805; DB 22; Length 27541;	
Dd	Best Local Similarity 53.0%; Pred. No. 2.9e-120; Indels 52; Gaps 11;	
Dd	Matches 1996; Conservative 0; Mismatches 1715;	
OY	4 GCAGATGTCCTCATCGACGCAGGCCGAAGAATTGGATGTCTGAGCAAGTTGC 63	
OY	21008 GCCGTCCCAGACCAGCGCCCCCGAAGCACGACCAATCGGATGTCGGATGACTGC 2106	
Dd	64 CGTGTCGCCGTGCGCGATGATCTGAGCGGGTCTTGAGAGCTTCTGAGGGCTCGCC 123	
Dd	21068 CGTAGACCCCGCGGGAGTGGCTCCGACGAGAACTGTGGCGGATCGCTTGAGAGAGTTC 2112	
OY	GACACCGTGGAGGAGTCCCGCGAAC---GCTGGATGAGCAGCGGTGTTGATCCC 180	
Dd	21128 GAGCGGATTCGGCTTCCCGCCGACCGCGGCTGGAGCGCCGAGGGCTCTAGACATCCG 2118	
OY	181 GACCCGATGCGCCCGGAGAAAGACGCCGTTACGCGCGCATCTTCTTGAAGCAGATAACC 240	
Dd	21188 GAACCCGACGCGCGCGGACCGACAATTAATCTCGTCCAGGGCGAATTCGCGGAGATCGCC 2124	
OY	241 TGCTTGAAGCTCTCTTCTTCCGATATCGCTCTCGAAGCGCTGCGGATGAGACCTCGCA 300	

Db	21248	GAGTTTCACCCGGGCTTCTTTCGGGATTCGCCCGCGAGGCGCTGTGCATGAGACCCGCGAG	21307
Qy	301	CATTGACTCTTGTCTGAGAGTGTGTCTGGAGGCGCTGGAGAACCGCCGCATTCGCTCCATCG	360
Db	21308	CAGCGGCTCTTCGTCTGAGAGACCGCGCTGGAGGCGCTTGAGACGCGCCGGCATTCGACCCGCTC	21367
Qy	361	GCGCTCGTCCGTACCGAAACGGGAGTTCATTATGGGATTCGGCCCTTCGATTATGAGGCC	420
Db	21368	GCGCAGCGCGCGCACCGCACCGGACCTTGTGTGGCGCTCAGTACACAGGACTTACCGCTTCC	21427
Qy	421	GCGCTGCGCGAAGGAGCGGCGTCCGAGATGACGCTCATGCGCGGCGTGGGAGCATG	480
Db	21428	GCGGTGCCCAACAGCGAGGGCTC-----GAGAGGCCAATGATACCGGACAGCTTC	21478
Qy	481	CCGACGCTCGAGAGCGGCGCGAATCTCGTATGCCCTCGGCGTGGAGAGGCGCTGTGTCTCGG	540
Db	21479	TCCAGTGTGCTGTCCGCGCGGATGTCCTTACTCTTTCGGCTTCGAGGCGCCCGCGCTCAG	21538
Qy	541	GTGGAATACGCTATTCCTGTCTCTGTGTGTGGCCGTTTATCTGGCCCTGTACAGCTTGGCC	600
Db	21539	CTCGACACCGGCT	21598
Qy	601	TCCGGGGGAATGCTCCACAGGCGCGCTGTGTGTGGGGTATCCGCTAATGTTGTTCGCGACAC	660
Db	21599	AACGGGGAGACTGCTGGCGCTTCGGCGGCGGTACAGCATATGTCACCCCGAATGTCG	21658
Qy	661	CTCGTGTGTCTCTCGAAGACCCGGGCGCTGTGGCCAGGACCGTGTGCTGCAAGGCAATTTTG	720
Db	21659	TTTCGTTCGGCTTCAACCGCGCAGCGCGCTCTGCGGAGACCGCGCTGTCAAGGGGTACGGG	21718
Qy	721	GCGGAGGCGGATGGGTTTCGAGCGAGCGAAGGGTCGCGCTGTGTCTTCAAGCGGCTTC	780
Db	21719	GACGCGCCCGCAGGGAGTGAACCTTCGCGAGGGGCTGGCGCTGTGTCTGTAGAGCGGCTG	21778
Qy	781	AGTGGACCCCGCGCGGACCGCGCATTCGGAATATGGCGGTATTTGAGAGATTCGCGATCAT	840
Db	21779	TCCGACCGCGCGCGCAACCGGACCAAGATCTCGCGTATTCGCGGACTTCGCGCTCAAC	21838
Qy	841	CACGACGGTGGAGACACGCGCTGTGACCGGTCCGACCGGAGCTCCCAAGAAATCGTGTG	900
Db	21839	CAGGACCGCGCTCTCAACCGCTTACCGGCACCCACCGGCCGCTCCAGAGCGCGTCAATC	21898
Qy	901	AAACGGGCGCTGTGGGAGCGCAGGCTGTGGCGCGCTTTCGGTGGTATATGTAGAGGACAC	960
Db	21899	CGCCAGGCGCTGTGGCAACTCCGCGGATGGGCCCGCGGACATCGATCTGAGGGGCGAC	21958
Qy	961	GGCACCGGCGACGAGCTTGTGTATCCCATGGAATTCAGACTTGAATCGGATACCGG	1020
Db	21959	GGCACCGGTACCGGCTCTCGGAGACCCATGAGGCGGAGGCGCTGTGTGTGTGTGTGTGT	22018
Qy	1021	CTCGGAGGAGATGTTCGCGACGCGCGCTCTGTATCGGGTCTGTGAAGACCAACTTTGGCAAT	1080
Db	22019	CAGGACCGCGCGCCCGGAAACGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	22078
Qy	1081	CCTGAGTATGTGCTGTGGGAGATCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1140
Db	22079	ACCCAGATATGGCATTCGGGATGTGCGACGATCATCAGTCTGTCCGCGCGCTTCAGGAAGG	22138
Qy	1141	CAGATTCCTGTGGGACCTTTCACAGCGCAGGCGCTGAACCCCGGATCTCATGGGATATCTT	1200
Db	22139	GTGTGTGCCAATCTCTGTGCATATGACCGGCGCTTCACCCAGCTGTGTCTTGTGGG	22198
Qy	1201	CGGCTGACCCGTCAACGCGCGCCCGGACACCGTGTGGGACTGTGAATATCGCGCAGGGG	1260
Db	22199	GCCATTCGGGCTGTCTCACCGAAACGACCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	22258
Qy	1261	GGGGTGAAGCTTGTTCGGCATGAGCGGAGACCAACCGCGCATGTGTGTGTGAAGAGCGCG	1320
Db	22259	GCGGCTCTCTCTTTCGGGATCAGGGGACCAACGTTCAACATCTCTGAAACGAGGCC	22318
Qy	1321	GCGGCGAGTTCACACGCGCGCGCGCGGAGGAGACCGGACAGTCTGTG-----GTGCTG	1374

Db	22319	GCAGACAGAGGCGCCCAAGCCCGGCCAACCGCGCGGAGAGGCGTGGTGGTCTCGTCTC	22373
QY	1375	TCGGCAGAGAGCCCGCTGACGCCCTTGATAGCAAGCGCGCGCGCGCTGCGACCATCTTGAG	1434
Db	22379	TCCGGCGCGGCGAGAGCGCGCTGCGCGGCCAGAGCGCGCGCGCTGCTCGCTTCGTGCGAG	22438
QY	1435	ACCTAACCTTCGCAAGTGTCTGAGGCGCATGTGTGGCGTTCAGTCTGAGCGACAGCGAGCGCG	1494
Db	22439	GAGCGGCGCCAGAGGCCACCTCACCGACCTGCGCCACTCCCTTGCCACCTGCGCGCGCGG	22498
QY	1495	ATGAGAGACCGGCGTCCGCGGTGCGGAGAGCGTCCAGAGGAGAGGCGCTGCGGCGACGCTTGAC	1554
Db	22499	CTGGAGAGCGCGCGCGCGCGCTGATCGCGCGCCAGACCGCGGACACCTTGACCCCGGCGCTGCGC	22558
QY	1555	GCTGCGGCGGAGGAGACAGAGCTGCGCGCGGTGCGGTGCGGAGTATGCGCGATTCTTCACGC	1614
Db	22559	GCCCTGTCCAGCGGCGGCGGCCAGCCCGGCGTGTGTCAAGGAGACCGCGGAG-----ACGC	22612
QY	1615	GGCAAGCTCGCCCTTTCTCTTTCACCCGGAACAGGGGGGCGACACGCTGGGACATGGGCGGTGG	1674
Db	22613	GGCGCGAGCCGCTTCGTCTGTTCACCCGGAACAGGGGAGCACAAGCGCCCGCGGATGGAGCGCGGA	22672
QY	1675	CTGTACGATGTATGTGTCCGCGTTCCCGCGAGGCGCTTGACCTGTGCGTGAAGCTGTTCAC	1734
Db	22673	CTCCACGACCGGCTTACCCGGGTTCGCGCGACGCGCTGAGAGAGTGTGGCCCGGCTCGAC	22732
QY	1735	CAGAGGCTCGACCGGCGCGCTCCGCGAGGTGATGTGGGCGGAAACGGCGACGCTGACGCGC	1794
Db	22733	GAGGAGACGGAGCCGGCGCGCTGCGGAGGTGTGTGTGCGCGCGCGACCTCCGCGAGGCGC	22792
QY	1795	GCGCTGCTCGACCAAGACAGGCTTACCCAGCGCGCGCTGTTCACCTTTCGATATATCGCTTC	1854
Db	22793	GCGCTCTCTGACCGGAGCCGGCTTACGCGCCACGCGCGCTTGTGCGCGTCAAGGTTCGCGCTG	22852
QY	1855	GCGCGCGTGTGCGGAGTGTGAGGGGTGTAGAGCCGGAATTGTGTGCGCGCGCCATATGACTGCT	1914
Db	22853	TTTCGCTGTGTGAGGTCTGTGGGCGCTGACCCCGGACTTACTGTGCGCGGCGACCTTCGTCGCG	22912
QY	1915	GAGCTGTGTGCTGCGCTGCGTGTGCGGAGCGGTGTCTCGCTTGAGAGACGCGGATTTCTTGATG	1974
Db	22913	GAACTGCGCGCGCGCGACGTCGCGCGCGGCGGTGTGTGTGCTGAGACGACGCTTGACTGTGTC	22972
QY	1975	GCTGTGCGCGCGCGCGCTGATGTGACAGGCGGCTTGCAGCGCGGCGCGGCGATGTGTTCGATCGAG	2034
Db	22973	GCGCGCGCGCGCGCGCTCATGTACAGGCGGCTTGCAGGAGGCGCGCGATGTGTGCGCCCTGAGG	23032
QY	2035	GCGCGCGGAGGCGCATGTGTGCTGTGCGGTGTGACGCGCGACGACGTCGCGTGTTCGATCGCGC	2094
Db	23033	GCGCGCGAGAGACGAGGTCTGTGCGCGCTCTGTGAGAGGAGCTTACGACCGGAGTTCGTCGCGC	23092
QY	2095	GCGGTTCACGCTCCGACCAAGGTGTTCATCGCGGCGCGCGGCGCAACCGGTGATGCGATC	2154
Db	23093	GCGGTTCACCGGCGCGCGCGGTGCGGTGTGTGTGCGCGGCGGTGTGAGAGGACGTGTCTCTTC	23152
QY	2155	GCGGCGGCGATGTGCTCGCGCGCGCGGCGCGCAACCAAGCGCTTCAAGTCTTCGATACGCTTC	2214
Db	23153	GCCGACCTTCTTTCGCGCGCGACGCGGCGCGCGGACCAACCGGCTCGGAGTGAACATGCTTTC	23212
QY	2215	CACCTCACCGGTCATGTGCGCGCGATGTGAGAGCGCTTGAGGCGGTGTGAGCGCAGTCCGCTGAGC	2274
Db	23213	CACCTGCGCGGTATGTAGCGGCATGTCTGACGACTTGTGCGCGCGTGCGCCCGGCGCTGAC	23272
QY	2275	TACCGGCGCGCGTGTGATGTCTGTGTCAAGCATCTGA---GCGGAGAGGCTTGTGACAGAC	2331
Db	23273	TACCAACCGCGACAGATCCCGTTGTGTGTGGAAGTTCAGCGGCGCGCTGTGACACCGCGGAA	23332
QY	2332	GAGGTGAGCTGCGCGGCGCTATTGGGTGTGCGCACGCGCGAGAGGTGTGCGCTTTCGCGAT	2391
Db	23333	CAGGTTCGACAGCGCGCGCTACTGTGTGTGAGGACGTCGCGCGCGGCTCGCTTCGCGCAG	23392
QY	2392	GGAATGAAGGCGCTGACGCG---GGCGGTTGCGGAGACCTGTGTGAGGTCGATCCGAA	2448
Db	23393	GGCATGTGATGTGTGCGCACCGGAGGAGGAGTTCACACCTTCTGTGAGCTTCGCGCGGAC	23453

OY	2449	TCGACGCTGCTCGGCGCTGAGGCGCTGCTGATGCGGAGG---CCGAGCGCGGCTGCTC	2505
Db	23453	GGCGTGTCAAGGTCATAGCCCGGAGAGGCTTACCGACCCGCTCCGACAGGACTGCTG	23512
OY	2506	GCATGTCGCGCGCTGAGCGTGAAGACCGGCGACCGTCTCTGAGGCGCTCGCGGCGTTC	2565
Db	23513	CCGACCCGCGCGCGGACCGGCGCGGAGAACCTGCTCTGATGCACCGGATCGCCGCGGCG	23572
OY	2566	TGGGCGCGTCCGCTGCGCTGCTCTCTGAGCCGCGCTTCCG---CTCAGGAGGAGCGGCG	2622
Db	23573	CACCGCGACCGCGCGCGCGTGCATGTGAGGCGGGTACTTCGCGACACAGCGCGCGCGCG	23632
OY	2623	GTGCGCGCTGCGCCAGTACCTCTTGGCAGCGGAGCGGCTACTGATTCAGACACGAAAGCGAC	2682
Db	23633	ACACAGCTGCGGACCTTAGCGGTTCCAGCGGACCGGATCTGCGGCCACACACCGCGCG	23692
OY	2683	GACCGCGCGCTGCGGCGACCGCGCTGCTCCGCGAGCGGGTCAAGACGAGTTCGAGAGGGG	2742
Db	23693	ACGAGCGCGCCACACAGCGCGGATCCGCGCTCGAGCGGAGTTCGTGGCGCGCGCTCGAGCGG	23752
OY	2743	GGCGCGGTCGCGCGCGCGGCGACCGGCGCGAGGCTCGGTCGACCATCGCGCGCGCGAGAC	2802
Db	23753	GACCAAGTCCGCGCGCTCTCGCGCGCTCCTTGAAGCTTGACGACCGCACCGTCAACCGGATG	23812
OY	2803	GGACCGCGGAGGAGTTCGAGCGCGCGCGGACCGGTCGTTCCGCTTCGAGTTCGATGAG	2862
Db	23813	GTCCCGCGGCTCACCGCGCTGAGCGCGCGCGCGCGGAG---CAGACCGAGCTGACCTC	23869
OY	2863	CCAGCGCGCTTATTCACCTCTGCTCTTGGGTCACGAGCGCGCGCGCCCTGTGTTGGCG	2922
Db	23870	TGGCGCTACCGCGCTCACCTGGAACCGCGCGCGCGCGCACCGACCGCGCGCTCCAC	23929
OY	2923	GAGGTTCGAGATTCGCGGCTTCGACGCGCGCGGCGCTCAGCTTCAATATATTCAGCTCGCGCTG	2982
Db	23930	GGCGGCTGCGCTCGTGTCTGTGTCCGACAGACACACAGAACCGTACGAGACGACGCGCG	23989
OY	2983	GGCATGTGTCCTCGACGACCTTGCAGGAAAGCCCAACCTTCGCTGCTCTCGAGGCGAG	3042
Db	23990	GCTTGGGCGACGCGGACGTCGAGACCGCGCTCGGCGCACACACCGTCCGCGCTGACGTCAC	24049
OY	3043	TGCGCGGCGGCGCATGTCGCGCGGTGGGAGAGGGGCGTGAAGCGGCTCGTGGTGGGCGAACCG	3102
Db	24050	ACCAACCGACGCGCGCGCGCGCTGCGCGCGCGGATCACCGGACCGCGG-----GG	24098
OY	3103	GTCATCGCCCTTTTCGCGGCGGAGCGTTTGCTACCAACGATCAACACGTCGCTGCGCTGCTG	3162
Db	24099	ACGAGGGCGCGTTTCAAGCGAGTGTGTGTCCTGTGCGCGCTCGGCACCGGAGACCGCGCG	24158
OY	3163	CTGCGCTCGGCTTCAAGGCGGCTTCTCGGCGATTCGAGCGCGCGCGCATGCGCCGTCCGCTACCTG	3222
Db	24159	ACCCCGGTGGCGCGCGCGCGCTTCAACCTTCACACACACCGCGCTTCAGGCGCTTCGCGAGCG	24218
OY	3223	ACGCGCATGTATTCGCGGCTTCGACAGAAATGACCAGCTTCAGCGCGGAGGAGCGGAGCTGATC	3282
Db	24219	CCGCGCATTCGAGCGCGCGCTGTGAGACGTACACCGCGGAGCGCTGTGCGCGCGCGCGCG	24278
OY	3283	CATCGCGCGACCGCGGCGGATTCGCTTCTGCGCGGATGCAGTGGCGGAGCACTGTGGAGCC	3342
Db	24279	AACAGGTTCACCGCGCGCGGACAGGCGCGCGCTCTGTGGCGCTGGCGCGCGCGCTGCGCTTG	24338
OY	3343	GAGGTTCATGCG--GAGGCGCGGCGACCGCGCGGAGAAAGCGCGCTACCTGAGAGTGCCTGGGCG	3400
Db	24339	AATCTGCGCGCGCGGTTTCGCGGCGACCTTCGACTCGCCCGACACCTTCGAGCGCGCAAGCGG	24398
OY	3401	TGCGGATATGTAGCGGATTCGCGCTTCGCGACCGGATTCGTTCGCGCGACGTCGCGCGCGTGGACCG	3460
Db	24399	CCCGCGGATTCGCGCGCGGATTCGCGCGCTACCGACCGCGAGACGCGGATGCGCGCTGCGCG	24458
OY	3461	GGCGCGAGGAGTACAGTCTGTCTCAACTCTGCTTTCGGCGGAGCTGATTCACAGATTT	3520
Db	24459	CTTCGCGGATTCCTCTCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	24518

Oy		3521	TCAATCTCCGCGCATGCAGACGGCGGATTGTGTGAAGTCTGGCAAGCGCAGCTTTACGGCG	3580
Dd		24519	CCGCTTCAGACCGCGCGCGCGGACACCGTCTTAATACCGCGGACCGCGGCGATTCGGG	24578
Oy		3581	ATAACGACCTCGGGAGCTCGCGCGCTTCCTGCGCAATCTCTCTTGCTGTGATCTCC	3640
Dd		24579	GCCACGTCGCCCGCGCGCTGTGGCCCCGGAGCGGCGCACCACTGCTGTCAACGACGCCG	24638
Oy		3641	GGGGGATGATGCTCTGACCGCGCGCGGAGTCCGTGCGCTCTTGGAGAAGTCTTCGGCG	3700
Dd		24639	GCGGCCCCGCGCGCGCGCGCGGACGCGCTCGCGCGCCGAACTGGAGAACTGGCGCCC	24698
Oy		3701	TGATCGCGGAGGCGGTTCACCCCTCCGCCCATCGCGAGCGCT	3743
Dd		24699	GGGTACCCTTCGCGCGCTGCGATCGCGCGGACCGGACCGCGCT	24741
<hr/>				
RESULT 8				
AAD17186				
ID	AAD17186	standard; DNA;	125401 BP.	
XX	AC	AAD17186;		
XX	DT	29-NOV-2001	(first entry)	
XX	DE	Streptomyces noursei	nystatin PKS gene cluster DNA.	
XX	KM	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;		
XX	KW	antifungal; antibiotic; ds.		
XX	OS	Streptomyces noursei.		
XX	Key	Location/Qualifiers		
FH	CDS	6337..34771	/tag= a	
FT		/product= "NysI complete protein"		
FT	CDS	34792..51099	/tag= b	
FT		/product= "NysJ protein"		
FT	CDS	51155..57355	/tag= c	
FT		/product= "NysK protein"		
FT	CDS	57503..58687	/tag= d	
FT		/product= "NysL protein"		
FT	CDS	complement (58786..58980)	/tag= e	
FT		/product= "NysM protein"		
FT	CDS	/note= "CDS does not include start codon"		
FT		complement (59045..60241)	/tag= f	
FT	CDS	/product= "NysN protein"		
FT		/note= "CDS does not include start codon"		
FT		complement (60238..61296)	/tag= g	
FT	CDS	/product= "NysO2 complete protein"		
FT		120628..121308	/tag= h	
FT		/product= "NysR4 (long) protein"		
XX	MO200159126-A2.			
XX	PD	16-AUG-2001.		
XX	PF	08-FEB-2001; 2001WO-GB00509.		
XX	PR	08-FEB-2000; 2000GB-0002840.		
XX	PR	10-APR-2000; 2000GB-0008786.		
XX	PR	14-APR-2000; 2000GB-0009387.		
XX	PA	(UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.		
XX	PA	(SUTP-) SINTER STIFTELSEN IND TEK FORSK.		
XX	PA	(ALPH-) ALPHARMA AS.		

PA (SINV-) SIVENT AS.  
PA (DZIE/) DZIEBEMSKA H.  
PA (ZOTC/) ZOTCHEV S B.  
PA (SEKU/) SEKUROVA O N.  
PA (FJAEV/) FJAEVLIK E.  
PA (BRAU/) BRAUTASET T.  
PA (STRO/) STROM A R.  
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;  
XX MPI: 2001-557614/62.  
DR P-PEDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,  
DR AAE10149, AAE10150.  
XX New nystatin polyketide synthase polynucleotides and polypeptides,  
PT useful as antibiotics and antifungals -  
XX  
PS Claim 1, Page 188-254, 266pp; English.  
CC The present invention relates to the cloning and sequencing of the gene  
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.  
CC The nystatin PKS is useful as antifungal antibiotics. The present  
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.  
XX  
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;  
Query Match 18.9%; Score 805; DB 22; Length 125401;  
Best Local Similarity 53.0%; Pred. No. 2,7e-120;  
Matches 1996; Conservative 0; Mismatches 1715; Indels 52; Gaps 11;  
QY 4 GCGGATCGTCCCATGAGCGCGCGCAAGATCCGATTGCGTCTGAGCGAGTTGC 63  
DB 21008 GCCGTCGCCACCGCGCGCCCGCAAGCGAGACCCGATCGATCTCGGCATGAGCTGC 21067  
QY 64 CGCTGCGCGGCGGCGTGCATGATCGAGCGGGTTCTGACGCTCTCGAGGGCTCGGCG 123  
DB 21068 CCTTACCCCGGCGGGGCTCGGCTCCCGGAGACCTGTGGCGGATCCCTGACGAGTTC 21127  
QY 124 GACACGCTGCGGCGAGTCCCGCGCGAAGC---GCTGGAGTACAGCAGCGTGTGATCC 180  
DB 21128 GACGCGATCTCGGCGTTCGCCCGCGACGCGGGTGGAGCGCGAGGCGCTTACGACCGG 21187  
QY 181 GACCCCGATCCCGCGGGAAGAGCGCCGTTACGCGCGCATCTTCTCTGAGCGATGCC 240  
DB 21188 GACCCCGACCGCGCGCGCGCATCTTCTCGAGGGCGGATCTCTGCGGACGTCGCC 21247  
QY 241 TGCCTCGACGCGCTCTTCTGCGCATCTGCGTCGCGAGCGCTGCGAGTACCGCTGCA 300  
DB 21248 GAGTTGACCGCGGCTTCTTTCGGAATCTGCGCGGAGGCGCTGTGAGTGAACCGCGAG 21307  
QY 301 CATCGACTTGTGAGGCTGTGCTGGAGGCGCTGGAGAGCGCGCGATGCTCCATCG 360  
DB 21308 CAGCGGCTCTGCTGGAAGACCGCTGGAGGGTTGACGACGCGCGGATTCGACCGCGTC 21367  
QY 361 GCGCTCTGTGTGAGGAAACGGAAGTGTTCATGCGGATGCGCCGTCGAAATAGAGCC 420  
DB 21368 GGCACGCGCGAGCGCGACCGCGCATCTTGTGTGGCGCGACGACTTACAGCTTCC 21427  
QY 421 GGGCTGCGCGAAGCGAGCGGGTCCGAGAGATGACGCTCATGGCGGGCTGGGGAGCAG 480  
DB 21428 GCGGTGCCCAACGCGAGGAGCTC-----GAAAGGCCAATGATCACCGGACGCTTC 21478  
QY 481 CCCAGGCTGAGAGCGGCGGAAATCTGTATGCTTGTGCGCTGCGAGGCGCTGTGTGCG 540  
DB 21479 TCCAGTGTCTGTTCGCGCGGGGTCTTACTTCTTGCGTGAAGGCGCCCGCGCTCAG 21538  
QY 541 GTGAGATCGGCTATTCTGTCTGTGCTGTGGCGCGCTTCATGTGCGCTGTAGAGCTTGGCG 600  
DB 21539 CTTCGACACCGCGCTCTCTCTCCCTGTGCGCGATGACCTGCGCTGCGACAGTCCCGCGC 21598  
QY 601 TCCGGGGAATGCTCAAGCGCGCTGTGGTGGGATTCGCTGATGTGTGCGGAGCACC 660

DB 21599 AACGGGAGAGCTCGCTGCGCTGGCCCTGGCCGCGCGCGTCAAGCATCATGTCCACCCCATGTCG 21658  
QY 661 CTCTGTGCTCTTCGAGAACCCCGGCGCTGCGCAGGAGCGGTCTGTCGAAGCATTTTCG 720  
DB 21659 TTCTGTGCGCTTCAGCGCGCAGCGCGCTCTCCCGAGGACGCGCGCTGTGCAAGCGTACCG 21718  
QY 721 GCGAGGCGGATGTGTTTTCGACGAGGCGAAGGATGCGCGCTGTGTGCTCAAGCGGCTC 780  
DB 21719 GACGCGCGCGAGGATACCTTCGCGAGGCGCTGTGCTGTGTCTGTGAGGCGCTG 21778  
QY 781 AGTGAAGCCCCGCGCGAGATCGGATGATATTTGCGGATTCGAGATCCCGCTCAAT 840  
DB 21779 TCCAGCGCCCGCGCAAGGCGCACAGGTGTCTGCGGTGATTCGCGGCTCCGCGTCAAC 21838  
QY 841 CACGACGCTGCGAGCAGCGGTGTGACCGGTGCGCAACGCGAGCTCCCAAGAAATGCTGTCG 900  
DB 21839 CAGGACGCGCTCCCAACGCGCTGACCGCACCCAAAGCGCGCTCCAGCAGCGCTCATC 21898  
QY 901 AAGCGGCGCTTGGCGGAGCAGCGCTGCGCGCGCTTCTTGGGTGTTATGTGAGGCAAC 960  
DB 21899 GCGCAGGCGCTGGCGCACTCGCGGTGCGCGCGCGCATGACGCTCTGAGAGGCGCAC 21958  
QY 961 GGCACGGCGCAGCGCTTGTGACCCCATTCGAAATCCAAAGCTTGAATGCGGTATACGCGC 1020  
DB 21959 GGCACGCGTACCGCTCTGCGGAGCCCATCGAGGCGAGCGCGCTCTGCGCACCTACGCGC 22018  
QY 1021 CTGCGCGAGATGTGCGCACCGCGCGCTGTGATTCGGGTGTGTAAGACCACTTGGCCAT 1080  
DB 22019 CAGGACCGCGCGCGCGCAAGCGCGCGCTGTGCTGTGCTGTGTAAGTCCAAATCGCGCAC 22078  
QY 1081 CTTGATGTATGTGCGGGGATCATGAGGCTGTGTAAGGTGCTGTGCTTCTTCAAGACGGG 1140  
DB 22079 ACCGATGTGATCTCGGCGGTGCGCAGCGTCAATCAAGCTGTGTCGCGCTTCAAGAGGC 22138  
QY 1141 CAGATTCTGTGCACTTCCACGCGCAGCGGTGAACCCCGGATCTCAATGAGGTGATCTT 1200  
DB 22139 GTGTGCGCCAGTCTCTGATGATGACGCGCGCTTCAACGATGATGATGATGATGATGATG 22198  
QY 1201 GCGCTGACCTGACGCGCGCGCGCGGACACCTGTGCGCGCATGGAATACCCCGGAGCGG 1260  
DB 22199 GCGATCGGCGTGTCTACCGAAGCGACCCCGTGGCGCGAGACCGCGCGCGCGCGCGCC 22258  
QY 1261 GGGGTGACTGCTTGTGAGTATGAGGCGGCAACGCGCGAGTGTGTGGAAGAGGCGCG 1320  
DB 22259 GCGTCTCTCTTGTGAGTATGAGGCGGCAACGCTGCAACATCTTGAACAGGCGCGCC 22318  
QY 1321 GCGCGAGCTGCAACCGCGCGCGCGCGCGAGCGACCGCGAGCTGCTG-----GTGCTG 1374  
DB 22319 GCGGACGAGGCG 22378  
QY 1375 TCGGCAAGAACCGCTGACCGCTGTGATGACAGCGCGCGCGCGCGCGCGCGCGCGCG 1434  
DB 22379 TCCGCGCGCGCGAGGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22438  
QY 1435 ACCTACCTTGTGCAATGTGTGCGGATGTGCGGCTTCACTGTGCGGAGCGAGCGCGG 1494  
DB 22439 GAGCGCGCGAGGCG 22498  
QY 1495 ATGAGCAACCGGCTTGTGAGTGTGCGCGCGAGCTGAGAGGAGGAGGAGGAGGAGGAGG 1554  
DB 22499 CTGGAACCG 22558  
QY 1555 GCTGCGCGCGAGGAGCAGAGCTGCGCGCGGTGTGCGCGAGTATGTGCGGATTCCTACGC 1614  
DB 22559 GCGCTGTGCGAGCGG-----AGCG 22612  
QY 1615 GCGAAGCTCGCTTCTTTCACCGGAGGAGGCGAGAGCGCTGAGGAGGAGGAGGAGGAGG 1674  
DB 22613 GGCAGGACCGCTTCTGTTCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 22672  
QY 1675 CTGTACGATGTATGTGCTGCGCGTTCGCGAGAGGCTTGAACCTGTGTGAGGAGGCTGTTCAC 1734



Dh 22673 CTCAACGCGCTACCCGGTGTTCGCCGACGCGCTGACGAGGTGTGAGCCCGGCTGCAC 22732  
Qy 1735 CAGAGCTCGACCCGCGCTCCGCGAGGTGATGTGTGGCCGAACCGGACGAGCTGCACGCC 1794  
Db 22733 GACGGAACCGAACCGGCGCTGCGAGAGTGTCTGTTCGCCGCGCTGCAGTCCGCGAGGCC 22792  
Qy 1795 GCGGTGTGACACGACGACCTTCAACCGGCGCGCTGTTCACCTTGAATATGCGCTC 1854  
Db 22793 GCGCTCTGACACGAGCCGAGCTACGCGCAGCCCGGCTGTTCGCCGTGAGTGTGCGCTG 22852  
Qy 1855 GCCCGCTGTGCGGTGTGTGGGTGTATAGCCGAGTGTGTGCGCGGCTATAGATCGGT 1914  
Db 22853 TTCCGCTGTGACGCTCTGAGGCTTGAACCGGCTACCTGTGCGCGGCTCTCGTGGC 22912  
Qy 1915 GAGCTGTGAGCTGTGCTGTGAGCGGCGCTGTTCCTGTGAGAGACGCGGTGTTCCTGTG 1974  
Db 22913 GAATCTGCGCGCGCGAGCTGTGCGCGGCTGTGTGCTGTGAGACGAGCTGTGACCTGTG 22972  
Qy 1975 GCTGCGCGGCGCGCTGTGATGACGCGCTGTGCGCGCGGCGGCGAGTGTGTGATCGAG 2034  
Db 22973 GCCCGCGCGCGCGCTGTGATGACGCGCTGTGCGCGCGGCGGCGAGTGTGTGCGCTGAG 23032  
Qy 2035 GCGCGGAGGCGCGATGTGTGCTGTGCGCGCTGTGCGCGCGCTGTGCGCGCTGTGCGCG 2094  
Db 23033 GCCCGGAGGACGAGGTCTGTGCGCGCTGTGAGAGGCGCTCACCGACCGGCTGTGCGCG 23092  
Qy 2095 GCGGTCAACGCTCCGAGACAGAGTGTGTATGCGCGCGCGCGGCGAACCGGCTATGCGATC 2154  
Db 23093 GCCGTCAACGCGCGCGGT 23152  
Qy 2155 GCGCGCGGATGTGCGCGCGCGCGCGCGCGGACCAAGGCGCTCACGCTGTGCGATGCTTC 2214  
Db 23153 GCCGACCTTTCGCGCGCGAGCGGCGCGCGGACCAAGCGGTGTGCGGTGTGCGCGCTTC 23212  
Qy 2215 CACTCAACGCTCAATGTGCGCGCGCTGTGAGAGCGGTGTGCGCGGTGTGCGCGGTGTGAGC 2274  
Db 23213 CACTCGCGCTGT 23272  
Qy 2275 TACCGGCGCGGT 2331  
Db 23273 TACCAACCGCGAGATTCGCTGT 23332  
Qy 2332 GAGGTGAGCTGTGCGCGCTATGT 2391  
Db 23333 CAGGTGCGGACGCGCGCTATGT 23392  
Qy 2392 GGAATGAGGCGCTGTGACGC---GCGGTGTGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGT 2448  
Db 23393 GGCATTCGATGT 23452  
Qy 2449 TCGACGCTGT 2505  
Db 23453 GCGGT 23512  
Qy 2506 GCATGT 2565  
Db 23513 CCGACCTGT 23572  
Qy 2566 TGGGCGGT 2622  
Db 23573 CACGCGGACGCGCGGT 23632  
Qy 2623 GTGCGGT 2682  
Db 23633 ACAGGCTGT 23692  
Qy 2683 GACGCGGCGGT 2742  
Db 23693 ACAGGCGCGCACGCGCGGT 23752  
Qy 2743 GCGCGGT 2802  
Db 23753 GACGACGT 23812

Qy 2803 GAGCGCGGAGAGGT 2862  
Db 23813 GTCCCGCGGT 23869  
Qy 2863 CAGGCGGT 2922  
Db 23870 TGGCGGT 23929  
Qy 2923 GAGGT 2982  
Db 23930 GCGCGGT 23989  
Qy 2983 GGCATGT 3042  
Db 23990 GCGT 24049  
Qy 3043 TGGCGCGGCGCATGT 3102  
Db 24050 ACCACCGACCGCGCGGT 24098  
Qy 3103 GTATGT 3162  
Db 24099 ACCAGGCGCGGT 24158  
Qy 3163 CTGCGT 3222  
Db 24159 ACCCGGT 24218  
Qy 3223 ACCGATGT 3282  
Db 24219 CCGGATGT 24278  
Qy 3283 CATGCGCGACCGCGCGGT 3342  
Db 24279 AACAGGT 24338  
Qy 3343 GAGGT 3400  
Db 24339 AACTGT 24398  
Qy 3401 TGGGT 3460  
Db 24399 CCGCGCGGT 24458  
Qy 3461 GCGCGGAGGAGT 3520  
Db 24459 CTTCCGCGGT 24518  
Qy 3521 TCAATGT 3580  
Db 24519 CCGCTGT 24578  
Qy 3581 ATTAACAGCTGT 3640  
Db 24579 GCAACGT 24638  
Qy 3641 GGGGAGT 3700  
Db 24639 GCG 24698  
Qy 3701 TGAATGT 3743  
Db 24699 GGGTACCTGT 24741

RESULT 9  
AA517367  
ID AA517367 standard; DNA; 33529 BP.  
XX AA517367;  
AC  
XX  
DT 12-MAR-2002 (first entry)



OY	1343	CGCCGAGACGACCGGGCAGAGCTGTGGTGTGCTGTCGCGAAGACGCGGGCTAGCCCTTGAGT	1402
Db	26627	AGCGGCGCGCTTCGCGCTATTTCGCGCTGGTCTGTTCGGATCGAAGACGAGCGCGGTAAATG	26686
OY	1403	CACAGCGCGCGCGCTGCGCGACCATCTGGAAGACTAACCTTCCTGCAAGTGTCTGGCGATG	1462
Db	26587	CGCAGCGCGGGCGGTGGCGGAAGTGGCTGGAAAGACACGAGGAGATGTGGGTGTGGACG	26746
OY	1463	TGGGCTTCACTGTGCGCGACACCGCGACCGCGATGTGAGACACCGGCTTCGCGTGGCGCA	1522
Db	26747	TGGTGCAGACCGCGCGCGCTGTGACCGACGCACTTCAGATCGCGAGCGGTGGTTCGCG	26806
OY	1523	CGTGAAGGAGAGGAGCTTCGCGCAGCCCTTGACGCTGCGCGCGCAGGACAGACGTCGCGG	1582
Db	26807	CGAGCGGCTCGCGGAGCTGTGTGAGGTCTTTCGCGCGCTGTGTGTGGGGCGCGCGAATGCGG	26866
OY	1583	GTCGCGTGTGCGCAGTATGTCCGCAATCTTCACGCGCAAGCTGCGCTTCTTCTTCAACCGAC	1642
Db	26867	CGGTGTGTGAGCGGAGACCGCGAA---GCGAGCGCGGAAGCTTGGCGGTGTGTTTCAACGGGCG	26922
OY	1643	AGGGAGCGCAGACGCTGGGCAATGGAGCGTGGGCGTGTACATGTATGATCCGCGTTCGCG	1702
Db	26924	AGGGCAGCGACCGGCTTCGGATGTGGGAAGGCTTTTACGAAGTACCCCGGTTCGCGT	26983
OY	1703	AGGCGTTTCAACTTGTGCGTGAAGGCTGTTTCAACGAGAGCTGACCGGCGGCTTCGCGAGG	1762
Db	26984	CGGCGTTTCAACGAGTGTGTGGAGGCGGTGGACCGGCACTTCGACCGGTGTGTGAGAGAG	27043
OY	1763	TGATGTGTGGCGCAACCGCGCACGCTGCAACGCGCGCTGCTTCACACAGACGTTTCAACC	1822
Db	27044	TGGTGTTCGCGCGCGCGGACGCGACGAGAAAGAGCGCAGCTGAGCGGACGGAACGAACTACCC	27103
OY	1823	AGCGGCGCTGTTCACCTTCGAATATCGGCTCGCGCGCGCTGTGAGCGGTCTGTGGGAGTGA	1882
Db	27104	AGCCCGGGCTGTGTGGCTGGAAGTGGCGCTGTACCGTCACTGTGGAGTCTGTGGGGCTGA	27163
OY	1883	AGCGGAGATTGTGTGCGCGCCATTACATCGGTGAGTGTGTGCTGCTGCTGTGGCGGCG	1942
Db	27164	AGCCCGCTGCGCTTCTGTGGGAGACTCGATPAGAGAGCTGACGCTGCGCACGTGGCGGTTG	27223
OY	1943	TGTTCTGCGTTTGAAGACGCGGTGTTCTGTGGGTGTGCGCGCGCGCGCTGTGATGCAAGCGCG	2002
Db	27224	TGCTGAGCTTGTGCGAAGCAGACGAAGCTAATGTGTGCGCGCGGTCTGCTGTATGCAAGGAT	27283
OY	2003	TGCCGCGCGCGGCGCGCATGTGTGATTCGAGGCGCGCGAGCGCAATGTGTGCTGTCGCG	2062
Db	27284	GCGAGGCGCGGGGAGCGATGTGTGTGTGTGGAGGCGTTCGAGCGCGAGGTGCACGCGGCGCG	27343
OY	2063	TGGGCGCGCAATGACGCTCGG-----TGTTCATTCGCGCGGTTCACGCTTCGGAACAG	2116
Db	27344	TGTGTGAGAGTGTGGGGCGCAAGGCGGCACTGAGCAATCGCGGGCTGAACGCGCCATGCA	27403
OY	2117	TGTCATTCGCGGCGCGCGGCAACCCGTGATGCGATTCGCGCGCGCGAATGCGCGCGCGG	2176
Db	27404	CGGTGTGTGAGCGGGGAAGAAAGCGCGGTGCTTCGCGGTGTGCGCGCGCGAGCTGTGAAGCGCAG	27463
OY	2177	GGGCGCGCAACCAAGCGCTTCAAGTCTTCGATGTGTTTCACTACCGTCTATGCGCCCGA	2236
Db	27464	GCCGCGCGACCGCGCTGT	27523
OY	2237	TGCTGTGAGAGGTTTGGGCGGTGTGGCGGAGTTCGAGTGTGCTGCGCGCGCGCGTGTGTGTC	2296
Db	27524	TGCTGTGAGAGTGTGTGGGAAGTGTGTGCGGAGTGTGTGATGCTGCGCGCGCGCACTGTGCGG	27583
OY	2297	TGCTCAGCAATCTGA---GCGGGAAGCTTGTGACAGACGAGTGTGAGCTGTGCGCGCTATT	2353
Db	27584	TGTGTGAGGCGGTGTGACGGGCGAAGCTTCGTGTGCGGAAGAAAGCGCTGTGTGTGCGCGAATCT	27643
OY	2354	GGGTGTGCGCAACCGCGGAGAGTGTGTGCTTTCGCGGATGTGAGTGTGAAGCGCGCTGACGCGG	2413
Db	27644	GGGTGTGAGCAAGTGTGTGAGGCGGTGTGTGCTTCTGTGACCGGAATGTGCGACGCTTTCGCGCGG	27703

Oy	2414	CCGGTGGGGGACCTTCGTGAGAGTGGAGTCCGAAATCGACGCTGCTGGGCTGGTGGCTCG	2473
Db	27704	CGGGGAGTGAACACATACGTGAGTGTGGGCGGAGTGGCGTCTGTGGCCGCTGGGGGCGG	27766
Oy	2474	CTTCGATGGCCGAGACGCCCCGCGCTGCTGCATC-----GTCCGCGCTGGGGCTGG	2527
Db	27764	GGTCCCTGGCCGGAGGGAGCCGAGGCACGTTTGTGGCGAGCCTGGCCGAGAGCAGAGG	27822
Oy	2528	ACGAGCCGAGGACCGTGGCTTCGAGAGGCTCGGCGGGCTCTGGGCGGTGGGTGGCTGTCT	2587
Db	27824	AAGAGCGCGGCGTGGGACGAGCGGTGGCGCAGTGCACGTGCAGGGGACGAGGTGAGCT	27883
Oy	2588	CTTGGGCGGGCTTCCTCCCTCAGGGGGGGGCGGGTGGCCGCTGCCACGTACCTTTGGC	2647
Db	27884	GAGGCCAAGTGTCTGTCCGGGCGGTGGGCGGCGGCGCCCTGTGAGGTGCGCAGAGTACGCTTTC	27943
Oy	2648	AGCGCGAGCGCTACTGGATTGGAATCGAACGAAAGCCGACGACCGCGCGCTGGCGACCGCGTGG	2707
Db	27944	AGCGGCGAGCCCTACTGTGCTGGAGAGCCCGAAGCGCGTACCGACGTGGGCTCGGCGGGCT	28003
Oy	2708	CTCCGGAGAGGGGTCAACGACGAGTGGAGAGAGGGGGCGCGGTGGCGCGGCGGACCGAG	2767
Db	28004	TGAGGAGGTGGGGGCATCCGCTCTCTGGAGCGGCGCAAGAGCTGGCCGACCGGCGAGCGCC	28065

CC	RESULT 10
XX	AAAF24892
ID	AAF24892 standard; DNA; 20394 BP.
AC	
XX	AAF24892;
DT	20-APR-2001 (first entry)
DE	Pimaricin biosynthesis associated polyketide synthase gene.
XX	
KW	polyketide synthase; oxidative modification; metabolite; antibiotic;
KW	anticancer; pimaricin; ss.
OS	Streptomyces natalensis.
XX	
FH	Key
FT	Location/Qualifiers
CDS	1..20394
FT	/*tag= a
FT	/product= "polyketide synthase"
XX	
PN	MO200077222-A1.
XX	
PD	21-DEC-2000.
XX	
PF	14-JUN-2000; 2000WO-EP6227.
XX	
PR	14-JUN-1999; 99EP-0201893.
XX	
PA	(STAM ) DSM NV.
XX	
PI	Martin JF, Aparicio JF, Collina AJ;
XX	
DR	WPJ.:2001-080693/09.
DR	P-Psdb; AAB31558.
XX	
PT	New polynucleotides encoding enzymes involved in the biosynthesis of
PT	pimaricin, useful for modifying the biosynthesis of pimaricin and in
XX	the synthesis of new compounds -
XX	
PS	Disclosure; Page 53-80; 116pp; English.
XX	
CC	The present sequence encodes a polyketide synthase which is associated
CC	with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC	is useful for the oxidative modification of a methyl group of a suitable
CC	compound, e.g. a bioactive compound including a secondary metabolite,
CC	antibiotics and anticancer agents. Recombinant cells comprising the
CC	gene are useful for the production of pimaricin. The polyketide synthase
CC	polynucleotide may be over expressed in Streptomyces, leading to an

CC increase in the biosynthesis of pimarinin, as a source of primers for  
XX amplification reaction and as probes.

Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

Query Match 18.0%; Score 766.6; DB 22; Length 20394;  
Best Local Similarity 53.6%; Pred. No. 4,1e-114;  
Matches 1044; Conservative 0; Mismatches 1549; Indels 46; Gaps 10;

```
QY 33 AGATCCGATTCGATTCGAGCCGATTCGCGTCTGCGCGGTGCGTATCGATCTGAG 92
DB AGAGCCCCCTCGGATTCGATTCGATTCGCGTCTGCGCGGTGCGTATCGATCTGAG 158
QY 93 CGGCTTCGAGAGCTTCCTGAGAGGCTCGCGGCAACCCGTCGGGCAAGTCCCGCGAAC- 151
DB 159 GAGAGCTGAGAGCTTCCTGAGAGGCTCGCGGCAACCCGTCGGGCAAGTCCCGCGAAC- 218
QY 152 -GCTGAGATGAGAGCTTCCTGAGAGGCTCGCGGCAACCCGTCGGGCAAGTCCCGCGAAC- 209
DB 219 CGGCTTCGAGAGCTTCCTGAGAGGCTCGCGGCAACCCGTCGGGCAAGTCCCGCGAAC- 278
QY 210 TACGCGCGATTCCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 269
DB 279 GCGCAGAGGCGGATTCCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 338
QY 270 GCGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 329
DB 339 CGGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 398
QY 330 GCGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 389
DB 399 AGCCGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 458
QY 390 CATCGGATTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 449
DB 459 TGTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 512
QY 450 GATTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 509
DB 513 GGTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 572
QY 510 TGCCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 569
DB 573 TGTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 632
QY 570 GCGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 629
DB 633 CGCCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 692
QY 630 TGGGATTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 689
DB 693 CGGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 752
QY 690 GCGCAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 749
DB 753 TGCCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 812
QY 750 AGGATTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 809
DB 813 GGTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 872
QY 810 ATTGATTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 869
DB 873 CCGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 932
QY 870 GCGCAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 929
DB 933 CGCCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 992
QY 930 GCGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 989
DB 993 CGCCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1052
```

```
QY 990 GGAATCCAGCTTCGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1049
DB 1053 CGAGGCGAGGCTTCGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1112
QY 1050 GATTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1109
DB 1113 ACTGAGGCTTCGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1172
QY 1110 GCTGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1169
DB 1173 GATGAAATTCGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1232
QY 1170 GCTGAAATTCGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1229
DB 1233 GCGCAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1292
QY 1230 GTGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1289
DB 1293 CTGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1352
QY 1290 CAACGCGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1340
DB 1353 CAACGCGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1412
QY 1341 GCGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1400
DB 1413 GCGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1472
QY 1401 TGCAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1460
DB 1473 TGCCAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1532
QY 1461 TGTGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1520
DB 1533 GGTGCGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1592
QY 1521 GACGTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1580
DB 1593 GCGCGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1652
QY 1581 CGGTGCGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1640
DB 1653 CGGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1706
QY 1641 ACAAGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1700
DB 1707 GCGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1766
QY 1701 CGAGGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1760
DB 1767 CAGGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1826
QY 1761 GGTGATTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1820
DB 1827 CATGCTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1886
QY 1821 CCAAGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1880
DB 1887 CCAAGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1946
QY 1881 AGAGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 1940
DB 1947 GCGACCGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 2006
QY 1941 CGTGTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 2000
DB 2007 CGTGTTCGAGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 2066
QY 2001 GCTGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 2060
DB 2067 GCTGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 2126
QY 2061 GGTGCGGAGGCTTCCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTC 2120
```

Db	2127	TCTGGCGGCGCACGAGGACCAAGTGGCCCTTGGCGGCGCGCAACGGAGCCCGATTCCACCTT	2186
Oy	2121	CATCGCGGGCGCCGGGCGAACCCGTGCGATGCGATCGCGGCGGCGATGCGCGCGCGCGGGC	2180
Db	2187	CATTTCGGGGCGACGGAACAGGCCGTCAACCGAGATGCGCGGCCCACTGGAGGGCACAGGCGCG	2246
Oy	2181	GCGAACCAAGCGCTTCCACGTTCTCGCATGCGTTCCATCACCGCTCATAGGCCCGATGCT	2240
Db	2247	CCGCAACCAAGCGGCTCGCGGTCAGCCAGCGCTTCCACTCCCGCACATGAGCAAGACTCT	2306
Oy	2241	GGAGGCGTTCCGGGCGGTGGCGGAGTCCGTGAGTCAACCGGCGGCGCGTGCATGTCGTGCT	2300
Db	2307	GGAGGACTTTCGGGCGCGTTCGCCCGCGGCTGACCTTCCACGCGCCCGCGATTCCTGTGT	2366
Oy	2301	CAGCAATCTGA---CGGGGAAGGCTTGCACAGACGAGGTGAGCTTCGCGGCGTATTGGGT	2357
Db	2367	GTCCACGGGTGACCGGCGCGCGCTCGCACCGAAGACGAACTGGCGTGGCCGACTACTGGGT	2426
Oy	2358	CGCCGACCGCGGAGAGAGTGTGTCCGCGGATGGAAGCGCTGCACCGGCGCG	2417
Db	2427	GGCGGAGGTTCGGCGGAACGTTCCGCTTGTGTGTCCGTGTGGCACCTTGTGAGCGAGGG	2486
Oy	2418	TGCGGGCACCTTCCGTGAGAGTCCGATCCGGAATTCAGCGGTGCTGGGCGTGGTGGCTGGCTG	2477
Db	2487	CGTACCAACCTTTCGTGGAGATGGACACGGCGGCGCTCTCACCCCATAGTTCAGAGACTTG	2546
Oy	2478	CATG---CCGAGCGCCCGGCGGCGGCGCTGCTGCATGTCGCGCGCTGGGCGTGCAGAGCC	2534
Db	2547	TCTGACCAACCTTCGGAAGAGCCGTTCTGTCCCTCTGTGCTGGGCAACCGGCGCGCCGGAAC	2606
Oy	2535	GGCGACCGTGTCTGAAGCGCGCTGGCGGCGCTTGGGCGGTCCGTGGCTGTGTCTCTGGGC	2594
Db	2607	CGTGGCTTCCACGAGGGCGTTCGCACCGCTTCTGTGACGAGTGTCCCGCTGCACCGGCTC	2666
Oy	2595	CGGCGCTTT---CCCTCAGGGGGGGCGGGGTGCGGTGCGGTGCAGTACCTTGGCAGCG	2651
Db	2667	CGCCTTCCCGGCGCGGCGCGGATCTCCCGCGGGACTGGCCCACTAGCGCTTTCAGCG	2726
Oy	2652	CGAGCGCTACTGATTCGACACGAAAGCCGACGACGCGGCGGTGCGACCGCGGTCTCC	2711
Db	2727	TCAGTGGTACTGGCTGTGACCCGCGCGGACACAGACAGGGGAGAGGCGCGCGCCGGAAGC	2786
Oy	2712	GGAGCGGGGTACGACGAGGTTCGAGGAGGGGGCGGGTCCGCGGCGGCGACCGGCGAG	2771
Db	2787	GGGCGAGG-----CCGGAATCTTGGGCGGCGGTGCAGACGAGACTTCAGAGGACTGT	2839
Oy	2772	CGCTCGGCTCGACCAATCCGCGCGCCGGAAGCGGAGCGCGGAGAAAGTTCAGAGCGCGCG	2831
Db	2840	CGGCGGTCTTGACCAATCGACGCGACGGAAGCGGACTCTCTCGGCGCTTCTGCGCCACCC	2899
Oy	2832	CGACCGTCCGTTCCGCGCTCGAGATGCATGAGCGCAGGCGGTGCTGATTCACCTCGTCTTCCG	2891
Db	2900	TCTCTCTCTGGCGAGGAGCG---CAGGATCCAGGCGCGCGGACCGCTTCACTAC	2956
Oy	2892	GGTCACGAGCGGCGCGCCCTGTGTCTGGGCGAGGTTCGAGATCCGCTTCGACCGCGCGGG	2951
Db	2957	GCACCCACTGGGCGCGCGGCAACCGCT-----CGGCGGAGCCCAACCGGCGCACTG	3011
Oy	2952	GCTCAGCTTCAATGATGTTCAGCTGGCGCTGGGCAATGTGTCCCAACACTTGCGGGAAA	3011
Db	3012	GCTCGTGTCTGCGCCGGAAGCGGACGACACCCGTGTGACCGGCGCGCTTCGTGAGCGC	3071
Oy	3012	GGCCAAACCTCCGCTGTGCTTCGAGGCGAGTGCAGCGGCGGCGCATCTGCGCGTGGGCGA	3071
Db	3072	GCTGACACACAGAGGCTGTGCACACCGAGTACGCAACTGCCCGCGCAACAGAGCCGA	3131
Oy	3072	GGGCGTGAACGGCGCTGTGTGTGGGCAACCGGTCAATCGCCCTTGGCGGGAGCGTTTGC	3131
Db	3132	CGCGCTGGGGCGACACCCCGTGTGAGCGGCGTGTCTGTGTGGCACTGCAGAGCGGCG	3191
Oy	3132	TACCCAGTCAACGTGGGCTGGCTGTGTGCTCGGCTTCAGGCGGCTTCGGCGAT	3191

ID	Accession	Gene	Protein	Length	Start	Stop	Score	E-value	Notes
Db	31192	CACCGGCTCCG	CGCCCTCGATACCGGGGCGGCGGCGGACCAACCAACGATGCTGGCGG	3251					
Qy	31192	CGAGGCGGCGG	CCATGCCCGCTCGCTACCTGACGCGCATGGTACCGCGCTCGACAGAAATAGC	3251					
Db	3252	CCCTGAGGAGG	CGGCGCATCCAGGACCGCGCTGTGTGCGTATCCCGCGCGCGCGCTCGCGCT	3311					
Qy	3252	CCGCGCTTACG	CGGGGGAGGGGGTGTGATTCATATGAGGCGACCGGCGGGGGTGGTCTCGC	3311					
Db	3312	CGACCGGACG	AGGCGGCTCAAGAGCCCTTACAGGACAGACATAGGAGGCGCTGGGCGGGGT	3371					
Qy	3312	CGCGGTCAGT	GAGGCGGACGACGCTGAGGAGCGGAGGTCATATCGACCGGCGGCGACGCGCGA	3371					
Db	3372	GGCGCGCCCTG	AGATGCCCGGACGAGCTGGGGGGGCGCTATATGACCTGCGCCGACAACTGGGA	3431					
Qy	3372	GAAACGCGCCT	ACTCTGAGTGTGCTGGGCGGTGCGGATATGTAGACGATTCGCCGCTCGACCG	3431					
Db	3432	CGGACGGGCG	GTCCTCCGCTGTGATGACACCTTCGCGGGGAGAGACAGCATGATCGCGCT	3491					
Qy	3432	GTTGTCGCG	CGACGTGGCGCGC 3452						
Db	3492	CCGCCCCCG	CGGGGTCTTCGC 3512						
RESULT 11									
AAT80413									
ID	AAT80413 standard; DNA; 43280 BP.								
XX	AAT80413;								
XX	27-FEB-1998 (first entry)								
DE	Tylactone synthase gene cluster.								
XX	Tylactone synthase gene cluster; tylG gene; multifunctional protein;								
KW	polyketide; tylactone synthesis; antibiotic; tylostin; ss.								
XX									
OS	Streptomyces fradiae.								
XX									
Key	Location/Qualifiers								
FT	816..14243								
FT	/tag= a								
FT	/transl_except= (pos: 816..818, aa: Met)								
FT	/note= "ORF1 encodes protein shown in AAW22601"								
FT	14351..19945								
FT	/tag= b								
FT	/transl_except= (pos: 14351..14353, aa: Met)								
FT	/note= "ORF2 encodes protein shown in AAW22602"								
FT	20010..31199								
FT	/tag= c								
FT	/transl_except= (pos: 20010..20012, aa: Met)								
FT	/note= "ORF3 encodes protein shown in AAW22603"								
FT	31232..36067								
FT	/tag= d								
FT	/note= "ORF4 encodes protein shown in AAW22604"								
FT	36249..41774								
FT	/tag= e								
FT	/note= "ORF5 encodes protein shown in AAW22605"								







CC Fragments involved in the biosynthesis of anamycins.

XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

Query Match 17.5%; Score 745.8; DB 19; Length 53789;  
Best Local Similarity 56.3%; Pred. No. 8.3e-111;  
Matches 1674; Conservative 0; Mismatches 1232; Indels 69; Gaps 12;

QY 28 GCCGAGATCCGATTCGATTCGAGCGAGTTGCCGTGCCCCGGGCGATGAT 87  
DB 41457 GCCGAGACCGCTCGCTCGTCGGGATGACCTGCGGTTCGGGGTGTCTCG 41516  
QY 88 CTGACCGGTTCTGACCGCTCTCGAGGGCTCGCGGACACCTGCGGGGAGATCCCCCG 147  
DB 41517 CCGGAGACCTGAGCTGCTGTGGCGGGGCTGAGCGCCCTTTCGACTTCCCGAC 41576  
QY 148 GAAC---GCTGGATATGACAGACGTCGTGATTCGCCGCCGATGCCCGGGAGAGC 204  
DB 41577 GACCGGGGCTGGGAGCTGAGCGGCTTTCGACCCGACCCGACCCCGGGACGTG 41636  
QY 205 CCGGTTACCGCGCATCTTCTCTGAGCGACGTAGCTGCTTCGACGCTCTTTCGAG 264  
DB 41637 TACACGACGACGCGGCTTCTCTGCGTGGCGCGGCTGTCGACGCGGGCTGTCGCG 41696  
QY 265 ATCTGCGCTTCGAGAGCGCTGCGGATGACCTTCGACATCGACTCTTGTGTGAGGTGCG 324  
DB 41697 ATCTGCGCTTCGAGAGCGCTTCGATGACCCGACGACGGGTGTCTGTGAGAGCTGCG 41756  
QY 325 TGGGAGCGCTGAGAGACCGCGATCGCTTCATCGCGCTGTCGATGCGGAAACGGGA 384  
DB 41757 TGGGAGCGCTTCGAGAGACCGCGGCTGACCGCTTTCGATGAGGACGACACGTCGCG 41816  
QY 385 GTGTTTCATGCGGATCGGCGCTTCGAAATGAGCGCGCTGCGCGACGCGCTCC 444  
DB 41817 GTGTTCTCGGCGCTTTCACCGAGGCTACGCGCGCGGCGG-----CGATCAG 41864  
QY 445 GAGAGATGAGAGCTCATGCGCGGCTGCGGAGAGATGCCAGGCTGAGACCGGGCGGAATC 504  
DB 41865 CCGGACCTGAGAGCGTTCCGCGGACATCGGGGCGGCTGACGCTGCGCGCGGCGG 41924  
QY 505 TCGTATGCGCTCGGAGCTGCGAGGCGCTGTGTCGCGGATGATACGCGCTATTCGTCGCG 564  
DB 41925 TCTGATGCTTCGCGGCTCGAGAGACCGGGGCTACCATGACACCGCGCTTCTGTCGTCG 41984  
QY 565 CTGATGCGCTTCATCTGCTGCTGACAGCTTGCCTGCGGGAGATGCTCCACGCGCTCG 624  
DB 41985 CTGATGCGCTTCATCTGCTGCTGACAGCTTGCCTGCGGGAGATGCTCCACGCGCTCG 42044  
QY 625 GCTGATGCGGATTCGCTGATGCTGTCGCGAGACCTGCTGCTGCTCGAAGACCGCG 684  
DB 42045 GCGGCGGGGCGACGCTGATGCTGCGAGCGCGGACCTTCTGCGCTTCGCGCGACGCG 42104  
QY 685 GCGCTGCGAGGAGCTGCTGCAAGGCAATTTTCGCGGAGGCGCGATGGTTCGACGCA 744  
DB 42105 GTGCTGCTGCGAGCGCGGCTGCAAGGCTTCTCTCGACCGCGGACCGCGCTGCG 42164  
QY 745 GCGGAGGCTGCGCGCTGCTGCTCAAGCGCTCACTGAGAGCGCGCGGACGCGCAT 804  
DB 42165 GCGGAGGCTGCGCGGCTGCTGCTGCTGCAAGCGCTTTCGCTGCGCGGACGCGCGCAC 42224  
QY 805 CCGATATGCGGCTGATTCGAGAGATCCGAGATCAATCAGACGCTGCGAGACGCGCTGCG 864  
DB 42225 CCGATTTCTGCGCGCTGCTGCGGCGAGCGGCTGCAAGATGAGCGCTTCGCAAGCGCTGCG 42284  
QY 865 ACCGTCGCAAGCGGAGCTCCGAAAGATCTGCTGAAACCGGCGCTTCGCGGACGCGAGCG 924  
DB 42285 ACCGTCGCAAGCGGAGCTTCGAGAGAGCGGCTGATCCGAGGCGCTTCGCGGCGCGCG 42344  
QY 925 TCGCGCGCTCTTCGCTGCTGATATGTCGAGGACACAGCGGACGCGGCTTCGTCGAC 984  
DB 42345 CTGCTGCGCTTCGATGCTGTCGAGGCTGTCGAGGCGCACGCGGACCGCGCTGCGCGAC 42404  
QY 985 CCGATGGAATCCAGCTCGATGAGATGCGGCTGCGGGGAGATGTCGCGACGCGG 1044

DB 42405 CCGATGGAAGCGGAGCGCTCTGCGACCTACCGGCGCGCTGAC-----CGGCGCG 42458  
QY 1045 CTGCTGATCGGCTCGTGAAGACCACTTTGGCCATCTTGATATGCTCGGGATCACT 1104  
DB 42459 CTGCTGCTGCGGCTCGCTCAATCGAATTCGCGGACACGAGGCGCGCGCGGGTTCGCG 42518  
QY 1105 GGGCTGCTGAAGGCTGCTTTCCTTCTGACGACGCGGACAGATTCCTGCGGACCTTCAGCG 1164  
DB 42519 GCGGTGATCAAGATGCTTCAGAGCCCTGCGGACCGCGGACGCGGACCTTCACGCTG 42578  
QY 1165 CAGGCGCTGAACCCCGGATCTCAAGGAGTATCTTCGCGTGAACCGTCAAGCGCGCGCG 1224  
DB 42579 GCGGAGCGGACCGGAGAGTGAATGCTGCGCGCGGCTGCGGATGACCGGACCG 42638  
QY 1225 ACACTGCGCGGACCTGAAATACCGCGAGCGGCGGGGTGAGCTGTTTCGAGATGAC 1284  
DB 42639 CCGGATGCGCGCGCGGATGATGCGCGCGCGGCGCGGGGTGTCGCGCTTCGAGATGAC 42698  
QY 1285 GCGACCAACGCGCACGCTGCTGAGAGAGCGCGCGCGGCGGACGCTGACACCGCGCGCG 1344  
DB 42699 GCGACCAACGCGCACGCTGATCTGAGAGAGCGCGCGCGGCGGACGCGGTCGAGAA 42758  
QY 1345 CCGGAGCGACCGG---CAGAGCTGCTGCTGCTGCTGCGCAAGACCGCGCTCAGCTGAT 1401  
DB 42759 CCGGATTCGAAGGAGCGGCTGCGCTGCTGCTGCGCGGACCGCACATCTTTGCGG 42818  
QY 1402 GCACAGCGCGCGCGGCTGCGGACCATCTGAGACCTAACCTTCGAGTGTTCGAGCAT 1461  
DB 42819 GCTCAGGCGCGCGCTGCGGAGATCTGCG---GCTCAGGCTGCTGCTGCGGCGCGG 42875  
QY 1462 GTGCGCTTCATCTGCGGACGACCGCGGACCGCATGAGACACCGGCTTCGCGTGGCGCG 1521  
DB 42876 CTGCGGACCGGCTGCTGCTGCGGCGGCGCTGCTGCTGCGGACCGCGGCTGCTGCGG 42935  
QY 1522 ACGTGAAGAGGAGGCTGCGGCGAGCGCTTCGAGCGCTGCGGCGGAGGACAGACGTCGCGC 1581  
DB 42936 GGAACGACAGAGAGCGGCTGCGGCGGCTGCGGCTGCGGCGGCGGCGGCGCGCGCG 42995  
QY 1582 GGTGCGGCTGCGGATGATCGGATCTCTACGCGGAGGAGCTTCGCTTCTTCACCGGA 1641  
DB 42996 GCGGCTGCGGCTGCGGCTGCGGCA---GACGCGGAGGCTGCTGCTTCCCGCGCG 43049  
QY 1642 CAGGAGGCGGACGCTTGGGATGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1701  
DB 43050 CAGGCTTTCGAGCGGCTGCGGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43109  
QY 1702 GAGGCTTCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1761  
DB 43110 ACGGCTTCGACGAGGCTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43169  
QY 1762 GTGATGCGGCGGACCGGCGGAG-----CGTCAAGCGCGGCTGCTGCTGCTGCTGCTG 1812  
DB 43170 CCGGCTGCGGAGCGCTGCTGCTGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43229  
QY 1813 GCTTCACCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1872  
DB 43230 GCTTCACCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43289  
QY 1873 TGGGCTGAGAGCGGAGTTCGCTGCGCGGACATGAGATGCTGAGCTGCTGCTGCTGCTG 1932  
DB 43280 TGGGCTGCTGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43349  
QY 1933 GTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1992  
DB 43350 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43409  
QY 1993 ATGCAAGGCTGCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2052  
DB 43410 ATGCAAGGCTGCGGCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43469  
QY 2053 GCTGCTGCGGCTGCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2112

Db 43470 GC-----CGACTGCTCGGCGAGCGCTGGAATCCGCCGCTCAACGCGCTTCG 43520  
Qy 2113 CAGGTGTCATTCGCGGCGCGCGGCAACCCCTGATCGATCGCGGCGCGATGCGCGC 2172  
Db 43521 GCGGTATCTCTTTCGCGGGAACGCGGATCGTCTCGCGCGCGCGCGCGATCGCGGAG 43580  
Qy 2173 CCGCGGCGCGCAACCAAGCGCTCCAGTCTCGATGCGTTCATCTCAACCGCTCATGCGC 2232  
Db 43581 CCGCGGCGCAACCAAGCAAGCTCAAGTTTCGACGCGGTTTCACTCCGCGGATGCGC 43640  
Qy 2233 CCGATGCTGAGGCGTTCGCGCGCTGTCGCGCGAGTCCGAGTCAACCGCGCGCTCATC 2292  
Db 43641 CCGATGCTGAGGCGTTCGCGCGCTGTCGCGCGAGTCCGAGTCAACCGCGCGCTCATC 43700  
Qy 2293 GTCCTGATCAGCACTTGA---GCGGAAAGCTTTCACAGAGAGTGAAGTCCGCGCGC 2349  
Db 43701 CCGGTGCTTCAACGTCGACCGCGCGGTTCCGCGCGAGCGCGCAACTGACCGAGCGCGC 43760  
Qy 2350 TATTGGGTGCGCACGCGCGAGAGGTGCGCTTCGCGATGAGTGAAGCGCGCTGCAC 2409  
Db 43761 TACTGGGCGCGACAGTGCAGCGCGCGCTGCGGTTCCGCGAGGCGCTGCGCGCGCAGC 43820  
Qy 2410 GCGGCGCGTGGCGGCGCTTGTGAGAGTCCGATCCGAATGACGCTGCTGCGCTGTG 2469  
Db 43821 GAGTCCGCGCGCTGCTGCTGTCGAGTCCGCGCGCGCGCGCGCGCTGACCGCTCGT- 43879  
Qy 2470 CTTGCTGATGCGCGAGCGCGCGCGCTGCTGCGATGTCGCGCGCGCGCGCTGAC 2529  
Db 43880 -----CGAGGAGAGCGCGCGAGTCACTGCTGCGCGCTGCGCGCGCGCGCGCGC 43931  
Qy 2530 GAGCGCGCGAGCGCTGCTGAGAGCGCTGCGCGCGCTGCGCGCGCTGCGCTGCTG 2589  
Db 43932 GAGGTCAACGCGCTGATCAACCGCGCGCTGCGCGAGTGTTCGCGCGCGGTTGCGGTGAT 43991  
Qy 2590 TGGGCGCGCTTCCCTCTCAAGGCGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 2649  
Db 43992 TGGCGCGCGCTGCTGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 44051  
Qy 2650 GCGGCGCTGCTGATGATGACGAGAAAGCGAGCGCGCGCTGCGCGCTGCTGCTG 2709  
Db 44052 CAGAGCACTTATGCTGCTGCGCGCGCGCGCGCGAGCGAGCGAGCGAGCGAGCGA 44110  
Qy 2710 CCGGAGCGCGCTGACGAGAGTGAAGAGGCGCGCGCTGCGCGCGCGCGAGCGCGC 2769  
Db 44111 GGTGCGCGCGAGCAACCGCGCTGCGCGCGCGCTGCGCGCGAGTGAAGCGCGCT 44170  
Qy 2770 AGCGCTGCGCTGACATCCGCGCGCGAGAGCGAGCGCGAGAGAGTGAAGCGCGC 2829  
Db 44171 GGTCTTCACTCG-----CGGCTGCTCAATTGAATCGACCGCGCTGCGCGAGCGTC 44225  
Qy 2830 GCGGCGCTGCGCTGCGCGCTGAGATGATGAGCGAGCGCGCTGCTGATCACTGCTGCT 2889  
Db 44226 ATCCGCGCGGTGCTGCTGCTGCGCGCGCGCGCTGCTGAGCTGCGCGCTGCGCGCG 44285  
Qy 2890 CCGGCTGACGAGCGCGCGCGCTGCTGCTGCGCGAGTGAAGTGAAGTGAAGCGCGC 2949  
Db 44286 GAGAGGCGCGCTGCGCGCTGCTGAGAACTGCTCACTGAGGCTGCGCTGCTGCTCC 44345  
Qy 2950 GGGCTCAGCTTCAATGATGATGCTGAGCTGCGCTGCG 2984  
Db 44346 GACCAAGCGCGCGCTGCGAGTCAAGTGTGCTGCG 44380

## RESULT 13

AA287283 standard; DNA, 15872 BP.

AA287283;

05-JUN-2000 (first entry)

S. venezuelae vep ORF 1, SEQ ID NO:1.

XX

KM Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
KM chronic obstructive pulmonary disease; respiratory inflammation;  
KM hypercholesterolemia; crop protection agent; de.

OS Streptomyces venezuelae ATCC15439.

FH Key Location/Qualifiers

FT CDS 20..13912

FT /tag= a /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"

FT CDS 14056..14151

FT /tag= b /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"

FT CDS 14167..15827

FT /tag= c /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"

PN WO20000620-A2.

PD 06-JAN-2000.

XX 25-JUN-1999; 99MO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MINI) UNIV MINNESOTA.

PI Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI; 2000-160679/14.

DR P-PSDB; AAV77177, AAV77178, AAV77199.

PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

XX synthesis of methymycin and pikromycin.

XX Example 3; Figure 23; 438bp; English.

XX The invention relates to an isolated and purified nucleic acid segment

XX comprising a desosamine biosynthetic gene cluster, a fragment or its

XX biologically active variant, where the nucleic acid sequence is not

XX derived from the eryl gene cluster of *Saccharopolyspora erythraea* or

XX streptomycetes antibiotics. The invention also relates to a macrolide

XX biosynthetic gene cluster, or fragments thereof. The macrolide

XX biosynthetic gene cluster encodes proteins which synthesize methymycin,

XX pikromycin, neomethymycin, narbomycin or a combination of these

XX compounds. Recombinant or augmented cells comprising the desosamine

XX and/or macrolide biosynthetic gene clusters are useful for the

XX production of biologically active macrolides. The macrolide biosynthetic

XX proteins are useful for synthesis of methymycin, pikromycin,

XX neomethymycin and narbomycin. The alternative termination of polyketide

XX synthesis may be useful to prepare novel antibiotics and

XX polyhydroxyalkanoate (PHA) monomers. The compounds produced by the

XX recombinant host cells are useful as biopolymers, e.g., in packaging or

XX biomedical applications, to engineer PHA monomer synthases or to prepare

XX biologically active agents, such as chemotherapeutics,

XX immunosuppressants, agents to treat asthma, chronic obstructive pulmonary

XX disease as well as other diseases involving respiratory inflammation,

XX cholesterol-lowering agents or macrolide-based antibiotics which are

XX active against a variety of organisms, e.g., bacteria, including

XX multi-drug resistant pneumococci and other respiratory pathogens, as well

XX as viral parasitic pathogens, or as crop protection agents (e.g.,

XX fungicides or insecticides) via expression of polyketides in plants. The

XX present sequence represents a Streptomyces venezuelae ATCC 15439 DNA

XX sequence, designated vep ORF 1 in the specification, which actually

XX contains 3 open reading frames, which encode proteins AAV77177-Y77178 and

XX AAV77199. The vep ORF 1 protein is defined in the specification as a PHA

XX monomer synthase.

XX

SQ Sequence 15872 BP, 2088 A; 6304 C; 5513 G; 1967 T; 0 other:

Query Match

17.4%; Score 742; DB 21; Length 15872;



[illegible]

RESULT 14	
AAI99682/c	
ID	AAI99682 standard; DNA; 4411529 BP.
XX	
AC	AAI99682;
XX	
DT	15-JAN-2002 (first entry)
XX	
DE	Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX	
KW	Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome; variation; epidemiology; patient treatment; epidemic monitoring; ds
OS	Mycobacterium tuberculosis.
XX	
PN	US6294328-BI.
XX	
PD	25-SEP-2001.
XX	
PF	24-JUN-1998; 98US-0103840.
XX	
PR	24-JUN-1998; 98US-0103840.
XX	

PA (GENO-) INST GENOMIC.RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX  
DR WPI; 2001-647261/74.  
XX  
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
XX 1551 and H37Rv differ --  
XX  
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
XX  
XX The invention relates to evaluating strain variation within and between  
XX different populations of the tuberculosis bacterial pathogen,  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
CC H37Rv (AA199682). The method is useful for evaluating strain variation ob  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
at [seqdata.uspto.gov/sequence.html?DocID=629438B1](http://seqdata.uspto.gov/sequence.html?DocID=629438B1).  
XX  
XX Sequence 4411529 BP; 758565 A; 1449983 C; 144602 G; 758379 T; 0 other;

Query Match	17.4%	Score 740.8	ID 22:	length 4411529:
Best Local Similarity	56.7%	Pred. No. 4,1e-110;		
Matches 1571:	Conservative 0;	Mismatches 1127;	Indels 74;	Gaps 8;
Qy 19	GAGCGCGACCGGAATCCGATTCGATCGTCGAGCGAGTTGCGCTGCGCCGGTGC	78		
Db 3297714	GAGCAACGCCGACAGAAACCGGTGCGGTGTGGGAATCGGCTCCGATTTCCGGGCGGC	32976550		
Qy 79	GTCATCGATCTGAGCCGGGCTTGTGACGCTCTCGAGGGGCTTCGGCGCACCCCTGCGGCGA	138		
Db 3297654	GTCGATGCCCCCGGACCGACCTCTGGGATGTGTTTGGCGGGGCGCGACGTGTGTCCGAG	32975959		
Qy 139	GTCGCCCGCCGAAAC---GCTGGAGTGCAGCAGCGTGTATCCCGACCCCGATGCCCG	195		
Db 3297594	TTTTCGACCGATCCGGGCTGTGGATGTTGAGGGGCTATACGATCCGATCCCGACGCTGAG	32975355		
Qy 196	GCGAGACGCCCGCTTACCGCGCATCTTCTTGAAGCAGCTAGCTCTTGACGCGCTCC	255		
Db 3297534	GCGAAACCTACACGCGATGGGGGCGATTCTTGAACACGCCACCGGCTTTGACGCGCGG	32974757		
Qy 256	TTCTTGGGCATCTTCGCTCGCGAAGCGCTGCGGATGGAACCTCGACATTCGACTTCTG	315		
Db 3297474	TTCTTCGGCATTCGCTCTTAATGAAATGTTGGCGATGATCTTCAGCAGCGGTGATCTG	32974155		
Qy 316	GAGGTGTCTGGAGGCGCTGAGAAACGCCGATCGCTCCATTCGCGCTCTGTGCTAGC	375		
Db 3297414	GAGGTTTCGTGGGAGGCGTTTGAGACAGCGGGGATTGACCCGTTGTGTTGCGGAGTTGC	32973555		
Qy 376	GAAACGGGAGTGTTCATTCGGGATTCGCGCCCTCCGAATATGAGCGCGGCTGCCGAAGC	435		
Db 3297354	GCGACCGGGGTGTACACCGGCAATTTTTCGCGCGAGTTACG-----GCAAT	32973101		
Qy 436	ACGCGTCCCGCAGAGATCGACGCTCATGTGCGGGCTGGGGACGATGCCACGTCGAGCG	495		
Db 3297309	AGGGATACCGGGGGGCTGCAAGGATACGGGTTGACCGGACGTTCATACACGTCGTGCTCG	32972550		
Qy 496	GCGCGAATCTTGATGCGCTTCGCGGCTGCGAGGGCGGTGTTCGCGGTGATACGGCTAT	555		
Db 3297249	GGAAGTGTGTCTGATACGATTTGGGGCTGCAAGGCGCCGCGGTGTCAAGTGATATACGCTTGT	32971900		
Qy 556	TTCGTCCTCGGTGAGCGGCTTCATCTGCGGCTGTTCAGAGCTTGGGCTCCGGGGAATGCTCC	615		
Db 3297189	TTCGTCGTCGTGGGCGATTCATGGGCGCATGTGTCGATTCGCTCCGGGGAATGTCAT	32971301		

QY 616 AGCGCCCTGGCTGGGGGATATCGCTGATGTTGTCGCGAGACACCTCGTGTGCTCG 675  
DB 3297129 TTGGGTGGGCGCGGGGATGACGGGTGATGGGCTGCGTCGATTTTGTGGGTTTTC 3297070  
QY 676 AAGACCCGGGCGCTGGCCAGGAGCCGTCGTCGACAGCATTTTCGCGGAGCCGATGG 735  
DB 3297069 CGGACGCGGGGTTTGGCTGCGGATGGGGGATGTAAAGGGTTTGGCGCGCGCGACGG 3297010  
QY 736 TTTCGACGAGCGGAGGGTGGCGCGCTCGTGTCTTCAGCGGCTCAATGAGAGCCCGCG 795  
DB 3297009 ACCGGTGGGGGAGAGGGCGCCGGGGTGGTGTGCTGAGCGGCTATCCGACCCCGGG 3296950  
QY 796 GACCGCATTCGGATTTGGCGGCTGATTCGAGGATCCGGGATCAACAGAGGTGCGAGC 855  
DB 3296949 TTGGGCGATTCGGTCTGCGGTGGGTGCGGGGAGCGGCTAACCGAGATGTGCTTC 3296890  
QY 856 AGCGGTGACCGGTGCGGACCGGAGCTCCCAAGAAATGTCGTGAAAAGCGGCGCTGCG 915  
DB 3296889 AATGGGCTGACCGGCGCCCAATGGGCTCGCGCAGCGAGTGTATCAAGTGGCGCTTGG 3296830  
QY 916 GACCGAGCTGCGCGCGCTCTTCCGTGGGTTATGTCGAGGACACAGCGACGCGGACG 975  
DB 3296829 AACCGGGGTTGTTCGCGCAGACCTGACGTCGTGAGGACACAGCGACCGCCACG 3296770  
QY 976 CTGGTGACCCCATCGAAATCCAGCTCTGATGCGGTAATCGGCTCGGGCGAGATGC 1035  
DB 3296769 TTGGGTGATTCCTGAGAGCGGACGAGCATTTGCTTCACCTTACGAGGAGAGTTCAG 3296710  
QY 1036 GCCACGCGCTGCTGATCGGTCGCTGAGAACCAACTTGGCCATCTGAGTATGCTCG 1095  
DB 3296709 GAGCAGCGGCTGTGGGTGGGCTCAATCAAGTGAATGGAATGGGACACACCGGCGCG 3296650  
QY 1096 GGGATCACTGGGCTCTGAGAGTCTTGTCTTCACGACGCGGACGATTTCTGCGAC 1155  
DB 3296649 GGGGTGCGCGGGGTGATCAAGATGTGACGAGGATGCTCATGCGGATGTCGCGACG 3296590  
QY 1156 CTCACGCGGAGGCGCTGAACCCCGGATCTCATGAGGGTGAATCTTCGCTGACGCTCAG 1215  
DB 3296589 TTGCATGTGATGATGACCTTCCCGGGGTGATTTGACACAGCGCGCGGTGCTGCTT 3296530  
QY 1216 CGCGCCCGGACACCTGCGCGGATCTGAAATACCGCGACGCGGCGGGGTGAGCTGTC 1275  
DB 3296529 ACCGAGCGCGAGGATGTCGCTGACGCGGCTCGCGCGCGGCGCGGTGCTCTT 3296470  
QY 1276 GGCATGAGCGGAGCCACAGCGGACGTGTGCTGAAAGAGCGCGCGCGAC----- 1328  
DB 3296469 GGGATCAACGCGACCAACGCGCATCTGATCTTGAAGAGGCCCGCTACACACCCCGCA 3296410  
QY 1329 -----GTGCACACCGCGCGCGCGGAGCGACCGGACAGCTGTGG 1369  
DB 3296409 GAGGCGCGGCTTGAAGCGAGCAAGACACCGGGGCGCGCGCTGATGTGCTGTGG 3296350  
QY 1370 TGTGTGCGCAAGACCGCGTCAAGCTTGATGACAGCGCGCGCGGCTGCGGACATC 1429  
DB 3296349 TGAATTCGCGAGATCGGCTGAGCGTGAACGCGCGGCGGCTGATATGCGCCACG 3296290  
QY 1440 TGGAGACTATCCCTTCGAGTGTCTGGGCGATGTGTGCGCTTCACTGTGCGACAGCGCA 1489  
DB 3296289 TGCAGGCCAACCCAGGGCTGATTCGATGTGAGGTGGGTGCTGTTTGGCAAT---CCCT 3296233  
QY 1490 GCGCATGAGAGCACCGGCTCGCGGTGGCGGACGTCGAGGGAGGGGCTGCGGGAGCGC 1549  
DB 3296232 CGGTGTTTGAAGCAACCGAGCGGTGTGTCGCGCAAGCGGTGAGCACTGATTTGCGGGC 3296173  
QY 1550 TGGAGCTGCGCGCGAGGACAGAGTGCCTCGGTGCGGTGCGGTGCGGATTCGCGCATTC 1609  
DB 3296172 TGGCTGGGCTGCGCGCGGCGAGCGGGTGCAGGTGCGGTGCGGTGCGGTGCGGTGCG 3296116  
QY 1610 CACGCGGCAAGCTGCGCTTTCTTTCACCGGACAGGGGCGGACAGCTGCGGCAATGGGC 1669  
DB 3296115 CGGTGGGCAAGAGCGGTGCTGCTTCTGCGGACAGGGCGCGGACGCAATCGGGAATGGGC 3296056  
QY 1670 GTGGGCTGACGATGATGTGCTCGGTTCGCGGAGGCGTTGACCTGTGCGGTGAGGCTGT 1729

DB 3296055 GCGAGTTGTACGGGAGTGTGCTCGGTGTCGCGAGGATTCGATGCGGTGCGGACAGT 3295996  
QY 1730 TCAACAGAGCTGACCGCGCGCTCCGAGGTGATGTGGCGCAACCGGACGCGTCG 1789  
DB 3295995 TGAACCGGATCTCGGTGCGCTGCGCGACGTTATTTGG-----GTGCGC 3295948  
QY 1790 ACGCGCGCTGCTGACAGACCTTCACCCGAGCGGCTGTTCACTTCGAAATATG 1849  
DB 3295947 ATGCGGATTTGCTGACAGACCAATTTGCTACGCCCGGCTGTTGCGGTGAGGTG 3295888  
QY 1850 CGCTCGCGCGCTGCGCGGTGCTGAGGATGAGAGCGGAGTGTGCTGCGCGCATAGCA 1909  
DB 3295887 CATGTTGCGGTGTTGCGGATTTGGGGTGTGCTTCCGACTTGTATGAGTATCTCG 3295828  
QY 1910 TCGGTAGCTGTGCTGCTGCGTGGCGGGCTGTTCTGCTTGAAGACCGGTTTC 1969  
DB 3295827 TTGAGAGCTGCGCGCGCGGACCGCGCGCGGTGTGTTGACGTTGCGGACCGCGCATGC 3295768  
QY 1970 TGTGTGCTGCGCGCGCGCTGATGACAGGCTGCGCGCGCGCGCGGCGGATGCTGTCGA 2029  
DB 3295767 TGTGTGCGCGCGCGCGCTGATGACAGGCTGCGCGCGCGCGCGGCGGATGCTGTCGA 3295708  
QY 2030 TCGAGCGCGCGGAGCGGATGTGCTGCTGCGGTGCGCGCGGACGAGCTGCTGTCGA 2089  
DB 3295707 TGTGCTGCGAGTACGAGT-----GAGCGGCTGCTGCGGTGAGGCTGCGGGA 3295657  
QY 2090 TCGCGCGGTCAACGCTTCGAGCAGAGTGTATCGCGCGCGCGCGGACACCTGTCATG 2149  
DB 3295656 TCGCTGCGATCAACGCGCGCGGATTCGGTGATCTCCGCGCGGACCGCGCAATG 3295597  
QY 2150 CGATTCGCGCGGATGCGCGCGCGCGCGCGCGGCGGACCAAGGCGCTCACTGCTGCGATG 2209  
DB 3295596 CGATTCGCGATGCGGTGCGCGCGGAGGTGCGCGGTCGACCAATGCGGTCTGCGATG 3295537  
QY 2210 CGTTCACTCAACGCTATGCGCGCGGATGCTGAGAGGCTGCGGCGTGTGCGCGAGTTCG 2269  
DB 3295536 CGTTCACTGCGCGCTTATGAGAGCGGATGCTGAGAGGTTGCGCGGTGCGCGCGCG 3295477  
QY 2270 TGAAGTACCGGCGCGCTGATGCTGCTGTCGACCAATCTGAGCGGAGAGCTTGCACAG 2329  
DB 3295476 TGCAGGACACGCGACCGGCTTGGGTGCTGTGTCGAACTGACCGGCGAGTTGCGCGC 3295417  
QY 2330 ACGAGTGAAGTCCCGCGGCTATTTGGGTGCGCGGCGGACGAGTGTGCGCTTTCGCG 2389  
DB 3295416 CTGATTCGCGGTGCGCGAGTACTGAGTGAACACGTTGCTGCGCGCGGTGCTTTCGCG 3295357  
QY 2390 ATGAGTGAAGCGCTGACCGCGCGGCTGAGGCGGACCTTGTGAGAGTTCGTCGAAAT 2449  
DB 3295356 ACATGCGCGCTCATTTGAGACCTTGGGGGAGACCACTTATGAGGCGCGCGGAGAA 3295297  
QY 2450 CGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2509  
DB 3295296 GTGGTTGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3295237  
QY 2510 CGTGGCGCGCTGCGGCTGAGGCGCGGCGACCGGCTGCTGAGGCGGCTGCGGCGGCTGCGG 2569  
DB 3295176 CCAACGCGGTGCGGTGAGTGTGCGGCGGTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 3295117  
QY 2627 CGTTCGCACTACCTTTCGAGCGGCGGACGCTACTGATGACACGAAAGCGACGACG 2686  
DB 3295116 AGCTGCCACGTAAGCTTTCAGGAGCGGCGGTTTTCGAGACCGCGGCGCGGATGCGC 3295057  
QY 2687 CGGCGCGTGGGAGACCGCGGCTGCTGCGGCGGCTGCGAGAGGTGAGAGGAGGCGGCG 2746  
DB 3295056 CGGCGGATGCGCGGCTGCGGCTGCGGCGGCGGACGACATGCTTGTGCGGTGCGGTG 3294997  
QY 2747 CGGTGCGGCGCG 2758

Db 3294996 TCGAGCGGCCCG 3294985

RESULT 15  
AAT68715  
ID AAT68715 standard; DNA, 15872 BP.

XX AAT68715;  
AC AAT68715;  
DT 01-SEP-1997 (first entry)

XX Streptomyces venezuelae polyketide synthase vep ORF1.  
DE Streptomyces venezuelae polyketide synthase vep ORF1.  
XX polyketide synthase; polyhydroxyalkanoate monomer synthase;  
KM polyhydroxybutyrate; biodegradable polymer; vep gene;  
KM metabolic engineering; ss.  
XX Streptomyces venezuelae.

OS  
XX  
XX Key Location/Qualifiers  
FH 20..13912  
FT CDS /\*tag= a  
FT CDS 14056..14136  
FT CDS /\*tag= b  
FT CDS 14148..15827  
FT CDS /\*tag= c

XX MO9722711-A1.  
XX 26-JUN-1997.  
XX 18-DEC-1996; 96WO-US20119.  
XX 19-DEC-1995; 95US-0008847.  
XX (MINU.) UNIV MINNESOTA.  
XX Sherman DH, Williams MD, Xue Y;  
PI WPI; 1997-341701/31.  
XX P-PSDB; AAM19629, AAM19630, AAM00918.  
XX  
XX Expression cassettes for production of polyhydroxyalkanoate(s) -  
PT provide wide range of biodegradable polymers for medical or  
PT industrial use  
XX  
XX  
XX Claim 54; Fig 23; 91pp; English.

XX Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide  
CC synthase (PKS) gene cluster encoding a polyene of 12 carbons (see  
CC also AAM19629-30 and AAM00918). It contains 5 PKS modules, with a 5'  
CC loading module and a 3' end domain. Each of the sequenced modules  
CC includes a keto-ACP, an acyltransferase, a dehydratase, a keto-  
CC reductase and an acyl carrier protein domain. The gene cluster was  
CC cloned using a heterologous hybridization strategy from a genomic  
CC DNA library. A novel expression cassette encoding the first module  
CC from the vep gene cluster and module 7 from the Streptomyces ty1P  
CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase  
CC activity and can be used for PHA prodn. in host (esp. insect) cells  
CC for use as a biodegradable polymer.  
XX  
XX Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

Query Match 17.2%; Score 735.6; DB 18; Length 15872;  
Best Local Similarity 55.6%; Pred. No. 3.8e-109;  
Matches 1590; Conservative 0; Mismatches 1219; Indels 53; Gaps 7;

QY 26 CAGCCGAGATCCGATGCGATCGTCCGAGCGAGTCCGCTCCGCGTGGCGTATCG 85  
Db 8490 CCGAGCGATCCGATCGATCGTCCGAGCGAGTCCGCTCCGCGTGGCGTATCG 8549  
QY 86 ATCTGACGGGCTTCTGACGCTCTCTGAGGCTCCGCGACACCGTGGCGAGTCCCG 145

Db 8550 CGCCGAGGACCTGTGCGGCTGTGCGCCGAGGAGCGAGCGGATCAGGATTCCTCCG 8609

QY 146 CCGAAC--GCTGGAGATGACAGACGCGTGTGATCCCGACCCCGATGCCCCGGAGAGA 202  
Db 8610 TCAACCGCGGCTGGAGCTTGAGAGACCTTACGACCCCGATTCGAGAGGCGACCA 8669

QY 203 CGCCCGTTACGCGCGATCTTTCTTGAGCGAGCTAGCTTCTTGACGCTCTCTTTCCG 262  
Db 8670 CGTACTGCGGAGGCGGCGGCTTCTTGAGAGGCGCCGCTGACTTCGACGCGCTTCTTCG 8729

QY 263 GCATCTGCGCTCCGAGACGCGCTGAGTGAACCTGACATCGACTTCTTGAGAGGT 322  
Db 8730 GCATCTGCGCGCGAGGCGCTTGTGATGACCCGACAGGCGCTCTCTGAGAGGT 8789

QY 323 GCTGGAGGCGCTGAGAGAGCGCGCATGCTCCATCGCGCTGTGCTGATCGAGAAACG 382  
Db 8790 CTTGGAGGCGCTGAGAGCGCGCGCATGACCCGCTCTGCTGCGGCGAGCGCGGTG 8849

QY 383 GAGTGTTCATCGGAGATCGGCGCTGCGATATGAGCGCGCTGCGCAAGCGAGCGCT 442  
Db 8850 GTGTCTACGTGGGCGCGCGCGACGCGCTGTAACGCTCCGATCCCGGCT-----GGTGC 8903

QY 443 CCGCAGAGATCGACGCTATGCGCGGCTGGGAGAGATGCCAGCTGAGCGGCGCGAA 502  
Db 8904 CCGAGGCGGTGAGAGGCGCTATCTGCTGACCGGAGCGCGCATGCTGATGTCGGCGCA 8963

QY 503 TCTGTATGCGCTCGGCGCTGAGAGGCGCTGTGCTGCGGCTGAGATACGCGCTATTCGCT 562  
Db 8964 TCTCTACGCGCTCGGCTGATGAGAGACGCTTCATGACGCTGAGAGCGGCTTCTCTCT 9023

QY 563 CGCTGTGCGCGCTTCATCTGCGCTGTCAGACTTGCCTCGGAGAAATGCTTCAACGCGCC 622  
Db 9024 CGCTGTGCGCGCTTCATCTGCGCGTACGCGGCTGCGGCGACCGGAGAGGCGGCTCGCG 9083

QY 623 TGGCTGTGCGGCTATGCTGATGCTGCTGCGGAGACCTCTGTGTGCTCTCAAGACC 682  
Db 9084 TGGCGGCGGCGGCTGCGGCTATGCGGATCGGATCGGCGGCTTCTCGGAGATTCCTCGGAG 9143

QY 683 GGGCGCTGCGCAGGAGACGCTGCTGCAAGGCAATTTTCGCGGAGAGCGGATGAGTTCGAG 742  
Db 9144 AGGGCGCTGCGCGCGCGAGCGCGCTGCAAGGCGTTCGCGCGCGCGCGCGACCGGCT 9203

QY 743 GAGGCGAGGAGTGCAGCGCTGCTGCTCAAGCGGCTCACTGAGACCGCGCGAGCGCG 802  
Db 9204 GGGCGAGGCGGCTGCGGCTGCTGCTGAGAGCGGCTGCTGAGACCGCGCGCGCGGCG 9263

QY 803 ATCGATATTGCGGCTGATTCGAGATCCGCGATCAATCAAGCGGTGCGAGACGCGCTC 862  
Db 9264 ACAAGGCTCTGCGCTGCTGACCGGCAACCGGCTCAACGAGACGCTGCTTCCACGCGG 9323

QY 863 TGAACCGGCGAGACGAGAGCTCCCAAGAAATGCTGAGAAACGAGGCGCTGCGGAGCGAG 922  
Db 9324 TGAACCGGCGAGACGAGAGCTCCCAAGAAATGCTGAGAAACGAGGCGCTGCGGAGCGAG 9383

QY 923 GCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 982  
Db 9384 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9443

QY 983 ACCCATCGAATTCAGAGCTGTAATGCGGTATACGCGCTTGGGCGAGATGTGCCAGC 1042  
Db 9444 ACCCATCGAATTCAGAGCTGTAATGCGGTATACGCGCTTGGGCGAGATGTGCCAGC 9503

QY 1043 CGCTGTGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1102  
Db 9504 CGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9563

QY 1103 CTGCGCTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1162  
Db 9564 GCGGCGCTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9623

QY 1163 CGAGGCGCTGAGACCGCGGATTCATGAGGCTATCTTGGCTGACCGCTGACGCGCGCC 1222  
Db 9624 CCGAGGCGCTGAGACCGCGGATTCATGAGGCTATCTTGGCTGACCGCTGACCGCTGCTGCG 9683

QY 1223 GGACACCGTGGCCCGGACTGGAATACGCCGCGAGCGGGGGTGAAGTCTGTTCCGCAAGA 1282  
 Db 9684 AGGTGCTTGGAGGAGACCGGCGCGGCGGAGCCGGAGGTGTCCTTCGGCGTCG 9743  
 QY 1283 GCGGAGACCAAGCGCGACTGGTGTGTGAAGAAGCGCGCGCGGCGAGCTGCAACCGCGCG 1342  
 Db 9744 GCGGACCAATATGCCATGTCTGCTCTGAAAGAGGACCGCGCGCGCGCGCGCGCGCG 9803  
 QY 1343 GCGCGGACGAC-----GGCAGAGCTGCTGG 1369  
 Db 9804 CCGGGAGAGGCGCGCGCGGCTCCCGCGCGGAGAGGGGAGGCGCGCTGGCTGGG 9863  
 QY 1370 TGTCTGCGGCAAGACCGGCTCAAGCTTGATGCAAGAGCGCGGCGGCTGGCGGCAATC 1429  
 Db 9864 TGTCTGCGGCAAGACCGGCTCAAGCTTGATGCAAGAGCGCGGCGGCTGGCGGCAATC 9923  
 QY 1430 TGAAGACCTACCTTTCGAGTGTCTGGGCGATGTGGCGTTCAAGTGTGCGAGCGGCA 1489  
 Db 9924 TCTCCCGACCGCGCGGCGCGCGCGGCTGATGCTTCTCTCCCTCCGCGCGCGCGCG 9983  
 QY 1490 GCGGATGAGACACCGGCTCGCGGTGCGCGCGAGCTGCAAGGAGGCGCTGGCGGCGCG 1549  
 Db 9984 CAGCTTTGACCAACCGCGCGCGGCTGATGCGGCTGGAAGGCGCGGCGCGCGCGCG 10043  
 QY 1550 TGGAGCGTGGCGCGGAGGAGACAGAGTGGCGCGGCGGCGGAGTATGCGGATTCCT 1609  
 Db 10044 TGGAGCGTGGCGCGGAGGAGACAGAGTGGCGCGGCGGCGGAGTATGCGGATTCCT 10100  
 QY 1610 CACGCGGACGCTCGCTTTCTCTTCAACCGGACAGAGGAGGCGGAGCGCTGGGCAATGAGC 1669  
 Db 10101 GGGAGCGGAGAGTGGCTTCTCTTCAACCGGAGGAGGAGGAGCGCGCGCGGAGGAGC 10160  
 QY 1670 GTGGCTGTAGAGTATGTTGCGGCTTCGCGAGGCGGCTTGAAGCTGTGAGGCTGT 1729  
 Db 10161 AGGAGCTGACG 10220  
 QY 1730 TCAACGAGAGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1789  
 Db 10221 TCGACCGCGCTGTGCG 10280  
 QY 1790 ACGCGCGCTGTGACGAGACAGAGCTTCAACCGGAGGCGGCGGCTTCACTTTCGAAATG 1849  
 Db 10281 AAGCGGCACTCTGAGACCGGAGCGGAGTACCCGAGCGCGCGCGCGCGCGCGCGCG 10340  
 QY 1850 CGCTGCGCGCTGTGCGGCTGTGCGGCTGTGAGACCGGAGTGTGCGCGCGCGCGCA 1909  
 Db 10341 CGCTGCGCGCTGTGCGGCTGTGAGACCGGAGTGTGCGCGCGCGCGCGCGCGCG 10400  
 QY 1910 TCGGAGACTGTGCTGCTGCGGCTGTGCGGCTGTGCTGCTGAGAGCGCGGCTTCC 1969  
 Db 10401 TGGGGAATGCG 10460  
 QY 1970 TGGTGGCTGCG 2029  
 Db 10461 TGGTGGCG 10520  
 QY 2030 TCGAGCGCGCGGAGCGGAGTGTGCTGCGGCTGTGCGGCGCGCGCGCGCGCGCG 2089  
 Db 10521 TCGCGGCGCGGAGCGGAGTGTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 10580  
 QY 2090 TCGCGGCGGTAACGCTCCGAGACGAGGTGTATGCGCGGCGCGGAGAGCGGCTGCG 2149  
 Db 10581 TCGCGGCGGTAACG 10640  
 QY 2150 CGATCGCGGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2209  
 Db 10641 AGCGCGCGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10700  
 QY 2210 CGTTCACTCAACCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2269  
 Db 10701 CTTTCACTCAACCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10760

QY 2270 TGAAGTACCGCGCGCGCTGCAATGCTCTGATGAGCAATGAGCGGAG---GCTTGA 2326  
 Db 10761 TGGCTTACCGGGAACCGGAGCTGAGGCTGCTCTTCAAGGTACAGCGGCGCGCGCGCG 10820  
 QY 2327 CAGAGAGTGAAGTTCGCGCGGCTATTGGGTGCGCAACCGCGGAGAGGTGTGCTTCC 2386  
 Db 10821 CCGGTGAATCAACGCGCGCGGACTACTGAGGTGCGCGAGTCCGAGACCGCGTGGCTTCC 10880  
 QY 2387 CGGATGAGTGAAGGCGCTGACCGCGCGCGGTGGGGGACCTTGTGTGAGGTGCGTCCA 2446  
 Db 10881 CGGAGCGGTGCGCGGAGCAACCGCTGAGGCGCGGAGCTTGTGTGAGGTGCGTCCA 10940  
 QY 2447 AATGAGCGTGTGCGCGGCTGAGTGTGCTGATGAGCGGAGCGCGCGCGCGCGCGCTG 2506  
 Db 10941 ACGGCGTGTGTGCGGAGTGTGAGAGAGTGTGAGAGAGACACCGTGGCGCTGCTG 11000  
 QY 2507 CATGTGCGCGGCTGTGAGCGAGCGCGGAGCGGCTGTGAGGCGCTGCGGCGCTT 2566  
 Db 11001 ---CGGAGATCAAGACCGCGGAGCGCGCGGAGCGGCTGTGAGGCGCTGCGGCGCTG 11057  
 QY 2567 GGGCGGTGTGAGGCTGTGCTCTGAGGCGCGGCTTCTCCCTGAGGGGGGCGGCGGCTG 2626  
 Db 11058 GGGCGGTGTGAGGCTGTGCTCTGAGGCGCGGCTTCTCCCTGAGGGGGGCGGCGGCTG 11115  
 QY 2627 CGTGGCCAGCTTGTGAGGCGCGGAGCGCTACTGAGTGAACGAGAAAGCGGAGCGAG 2686  
 Db 11116 GAGGCGGCGCGGAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11175  
 QY 2687 CGGCGGTGTGAGGCGCGGCTGTGCGGAGCGGCTCAAGAGAGTGAAGAGGCGGCG 2746  
 Db 11176 CGGCTACTGAGTGTGCGCGGCGCGCGCGCGCGCGGAGCACCGAGCACTGTATGATCGG 11235  
 QY 2747 CGGTGCGCGGCGGAGACCGGCGGAGCGCTGCGGCTGACATCCGCGCGCGGAGCGGAGC 2806  
 Db 11236 CTGGAGACCGGCTGTGCGGCTGTGAGCGGCGGCGCGCGGAGCGCGGCGGCTGTGAT 11295  
 QY 2807 GCCGGAGAAAGTGAAGCGCGCGGAGCGGCTGCGGCTTCCGCG 2848  
 Db 11296 CCAACCGGAGCGCGCGGCGGCTGTGCGGAGCGTGTCCGCGGCGAGCG 11337

Search completed: October 3, 2003, 12:27:59  
 Job time : 1144.88 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:41 ; Search time 15260.8 seconds

(without alignments)  
11435.853 Million cell updates/sec

Title: US-10-014-717-1\_COPY\_7610\_11875

Perfect score: 4266

Sequence: 1 gtggcgagatcgctccatcga.....gggaatcattgcctatga 4266

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_bg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_stb:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_stb:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pln:\*  
35: em\_hlg\_rdi:\*  
36: em\_hlg\_mam:\*  
37: em\_hlg\_vrt:\*  
38: em\_gy:\*  
39: em\_hlgo\_hum:\*  
40: em\_hlgo\_mus:\*  
41: em\_hlgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4266	100.0	68750	1 AF210843	AF210843 Sorangium
2	4266	100.0	68750	6 AR193029	AR193029 Sequence
3	4266	100.0	68750	6 AR199551	AR199551 Sequence
4	4266	100.0	68750	6 AR199559	AR199559 Sequence
5	4266	100.0	68750	6 AR199567	AR199567 Sequence
6	4266	100.0	68750	6 AR201097	AR201097 Sequence
7	4266	100.0	68750	6 AR208671	AR208671 Sequence
8	4187.6	98.2	58733	1 AF211789	AF211789 Sorangium
9	4187.6	98.2	71989	6 AR172664	AR172664 Sequence
10	1287.4	30.2	16124	6 AX024277	AX024277 Sequence
11	928.8	21.8	10910	6 AX024276	AX024276 Sequence
12	859	20.1	48000	1 POL505006	POL505006 Sorangium
13	844.4	19.8	49736	1 AF319998	AF319998 Stigmatel
14	817.2	19.2	66808	1 SAU421825	SAU421825 Stigmatel
15	814.2	19.1	3978	6 AR266876	AR266876 Sequence
16	814.2	19.1	12249	6 AR266931	AR266931 Sequence
17	814.2	19.1	18331	6 AR266948	AR266948 Sequence
18	812	19.0	42603	1 AF188287	AF188287 Stigmatel
19	805	18.9	27541	6 AX211706	AX211706 Sequence
20	805	18.9	123580	1 AF263912	AF263912 Streptomy
21	805	18.9	125401	6 AX211739	AX211739 Sequence
22	803.2	18.8	33529	6 AR166425	AR166425 Sequence
23	790.6	18.5	290850	1 SC0939127	SC0939127 Streptomy
24	784	18.4	84985	1 SNA278573	SNA278573 Streptomy
25	772	18.1	113193	1 AF357202	AF357202 Streptomy
26	772	18.1	113193	6 AX703543	AX703543 Sequence
27	770.8	18.1	5355	6 AX697895	AX697895 Sequence
28	770.8	18.1	60196	6 AX697977	AX697977 Sequence
29	766.6	18.0	20394	1 SNA13222	AX13222 Streptomy
30	766.6	18.0	20394	6 AX067996	AX067996 Sequence
31	766.4	18.0	39314	1 SGR300302	SGR300302 Streptomy
32	765.8	18.0	69644	1 AY179507	AY179507 Streptomy
33	761.2	17.8	104326	1 AB070940	AB070940 Streptomy
34	761.2	17.8	281450	1 AP005032	AP005032 Streptomy
35	755.2	17.7	43280	1 SFU78289	U78289 Streptomyce
36	745.8	17.5	53784	1 AMM223012	AMM223012 Amycolato
37	745.8	17.5	53789	6 A69720	A69720 Sequence 3
38	745.8	17.5	90445	1 AF040570	AF040570 Amycolato
39	741.4	17.4	318050	1 EX248344	EX248344 Mycobacte
40	740.8	17.4	16096	1 AE007124	AE007124 Mycobacte
41	740.8	17.4	24292	1 MTCY24G1	Z8388 Mycobacteri
42	732.8	17.2	78210	1 AB070949	AB070949 Streptomy
43	732.8	17.2	300425	1 AP005022	AP005022 Streptomy
44	726.2	17.0	65140	6 AX211705	AX211705 Sequence
45	725	17.0	49377	6 I88042	I88042 Sequence 1

## ALIGNMENTS

RESULT 1  
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000  
DEFINITION Sorangium cellulosum strain So ce90 epochlone biosynthesis gene  
cluster, complete sequence.  
ACCESSION AF210843  
VERSION AF210843.1 GI:6724237  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
Polyangium cellulosum  
Polyangium cellulosum  
Bacteria: Proteobacteria: Deltaproteobacteria: Myxococcales;  
Sorangineae; Polyangiaceae; Polyangium.  
1 (bases 1 to 68750)  
Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milamow,M.,  
Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,

TITLE  
The biosynthetic gene cluster for the microtubule-stabilizing  
agents epothilones A and B from *Sorangium cellulosum* So ce90  
JOURNAL  
Chem. Biol. 7 (2), 97-109 (2000)  
MEDLINE  
20130945  
PUBMED  
10662695  
REFERENCE  
2 (bases 1 to 68750)  
AUTHORS  
Molnar, I.  
TITLE  
Direct Submission  
Submitted (03-DEC-1999) Natural Product Genetics, Novartis  
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box  
12257, Research Triangle Park, NC 27709, USA  
FEATURES  
source  
1. 68750  
/organism="Polysangium cellulosum"  
/mol\_type="genomic DNA"  
/strain="So ce90"  
/db\_xref="taxon:56"  
<1..1826  
/note="ORF1"  
/codon\_start=3  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26904.1"  
/db\_xref="GI:6724238"  
/translation="ASIDLPARATSAVLDHGRTAERHYLAERAGIEDRALREH  
LRIORGSPFHCIGDLTVELAHDOPLASISPHARSLRHPDWTSDAMLYGPAV  
RLAARGAPGLREYEERERARTAOEARLWLAAPCFAPDLPRFEDNAGLELP  
MSPEVAERRLRASVATPELCAALWLTGAGWSGPYEMLPENLLGFIGLPT  
ALAAASPTGSEALRGARLPASWEVSSKSGOGLPEALMERLRTTYRAMGNAD  
LSRPEREALIAEVERLRAOPAPPAAGLAIVAGVSSGRLGLVTDGALVSGND  
TYMFORISPIVLTACTDPPEELAPLSOMLVANANAGTISKLTGESSPLIWAR  
QARPSLVHARFMAVNOAMPDPERCAPVYVQSTIMEFHPIPRCLHEPAGSFS  
LACDEHLWCESLISAGRLMWRPHRPAERFAYLGHPIAATWYPSLTINATHV  
MADPRRALIGYDKRTGVEPIVLAETRHPAPVASEDRDIPALTGQPSRDWVHEIR  
SASTVADVQOLMDRPMVLRNCLFTTNDRIITLARS"  
complement(1900..3171)  
CDS  
/note="ORF2; similar to hypotheetical proteins from  
mycobacterium tuberculosis and Streptomyces coelicolor and  
to Dd-peptidases"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26905.1"  
/db\_xref="GI:6724239"  
/translation="MGALISVAAPGALGAEERQPGDAGALAPAREVMAAEVA  
AQMPGAVLVARGDVIDAVAGTELGGASPMWRDITPRIASMTKAVATAVMMVVE  
EGKLDLDSVDWMLPELANKYLAIDGPIDETPAERPIITVADLMTFTMGISFDA  
SSPIORALDELGLVNAQPVMTPHGDEWIRLGTLPKHOGAOMVMYTGSLVQGL  
VGRADOGFATVREIRILALGMROTDHFVPRDKLARFAGCGYFTDDEORTGKTRMD  
GASAVASBPAPPSGAAGIVSTVYDYLIFARNLMNGVHIEGRLLISAASVREKTYDL  
TAAQKASFFPGFFETHGMVGMAVTVAPDAVSEVPGYGMDDGGGTISWINDPGRRL  
IGIWMQASGLFESGALREFMRSVYVATSA"  
3415..5556  
/note="ORF3; similar to Na/H and drug/H antiporters"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26906.1"  
/db\_xref="GI:6724240"  
/translation="MHGLTEROVLLSTVLTALILTVARASGELARLRPOEVLGELFG  
GVILGPSVVGALPGFHRALPOEPANGVLISGISMGLALLMLIGIEDVAILKEKA  
RGALSALGATIPPLAAGAFSALVUDRLPBGSLPGITYLSTANVIAKVILIEKES  
RNSVACVTLAAGVSEVAAMVAVMTSSYSGSPALAVASALMSGLIFMLVIGKR  
LTHLAMRVADATRVSKGVSLVLTFLAALTLTGAPLGLGAFGLVNLASPT  
NRPLDGVTLVAGLFAVPFVYLAGRVVSOQLRPLAAGTALLALLETAKVSPAA  
LGARLGLGSEBALVAAGLKKGGTDLVAVLGVGLISBAATMVAVVAVLVTVTA  
SPALLIWEKRAPPOESARLEREERARAYIPGVERTLPIVAHALPGPATDYES  
IYASRKRGCTVDITELSYEQAPGGRAPAGASRLAALGARLVGTRQGRRLRGS  
IOAILASRDHDLVIGARSAPARAGMSFGRLQDAIVQAESNLVAVVQDPDAERAS  
ARILVPIIGLESPALADLAHVLAAMDPAEVLVLSAQTDPGAIVMVRDPSRVA  
ARSVDEAVFRGRILGVRSRVHVGASDETTELARAPDLVLVLCYDHDGPIGLRL  
YIGSTVESVVRSRVVALVHAGTREGVR"

CDS  
complement(5612..5992)  
/note="ORF4"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26907.1"  
/db\_xref="GI:6724241"  
/translation="MKKIDGTGRCAIAEGVIRPGSNGDEPOMTSHERTCLINASDRDA  
QVALTVFSDDPAPGPKRYVTPARTRHVRNNDLTPPEIPRDIDYASVISEDPIV  
QHTRLDSROAENALSTIAYTDRE"  
6226..6675  
/note="ORF5"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAF26908.1"  
/db\_xref="GI:6724242"  
/translation="MKHVDGTGRFRGRRTIGTLGLASNALACGSPSEKTVQGTRLAP  
GADARVTADVDPDAATTRLAVDVVHLSPERLEAGSERFVVMORRSPSPRRVGVLD  
YNADSRCKLAETTPVYANFELLITAEKSSPQSPSSAAVIGTISVC"  
7610..11875  
/gene="epoa"  
/gene="11875"  
/gene="epoa"  
/note="BPOS A"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAF26919.1"  
/db\_xref="GI:6724253"  
/translation="MADPRIRAAEDPIAVGASCRLPGLGVIDLSGFWTLLEGSDTV  
GVAPREDAAMDPDPAARGKTPVRASPLSDVACPDASFPGISPEALRMDPRH  
LLEYCHCALNMAIAPBALVGTGTGVTGIGSPREYEAALQATRASEIDHAGLGT  
PSVAGASTYALGRCVAVDTYSSVLAHACQSLRSGESTALAGVSLMLSP  
STLWLSRTPRALDGRCKASASADGREGGAVVVKLSGARLADGDIILVIRG  
SAIHNDGSSGLTVPNGSSOEIVLKRALADGCAASGVYEAAGTTLGDDPLEIQ  
LNAVIGDRVATPLTIGSVKTNLGHPEVAGTIGLKVLSVLSHGOIPALHMOALN  
PRISGDLRLTVFARPTPMPTMPRAGVSGFMSGTHNAVULEAPAACTCPAPE  
RPAELVTSATTSALNDQAARLADHLETYSQCLGVASPLATTTRSMERLAAAT  
SREGIRALNDAAQOQTSPGAVERIADSSROKLAFLFTGQAQTLGKGRGLYDWSAF  
REAPDLVRLFNQIEDRLREVMMAEPASVLAALDQTAFTQPLFTFEVLAALMRS  
MGVEPELVAGHSIGELVAVAGVSEDAVFLVAARGLMQLPAGAMVSTEAPEA  
DVAAVAPHAASVSIAAVNAPDOVVIAGAGOPVHAIAAMARGARTALHVSHPFS  
PLMAPLEAPRVAESVSRPSIVIVNSGKCTDDEVSSRGVWRHARRVPRADG  
VKALHAAAGTFVEYKPTSLGLVYKAPMPARALLASSAGDEPATVLEAGLW  
AVGLVSWAGLPFSGGRVPLPTIPWQEKRTWITKADARGRARRAGACHDEYEG  
GAVRGDRSRRLHPPESEGRREKVEAAGRPRLTIDEGVADHLVLRTERAPG  
LGEVEIADAGLSEFNDVQALMGVPPDLPCKPPLPLGECAGRIVAVEGVAGLV  
VGOPYIALSAGAFTHVTTSAALVLPRLQALSAIEAAMPVAYLTAWVLDRIARLP  
GERVYIHAATGVGLAAVOMOHGAEVHATGPEKAYLESIGVRYVSDRSRDFV  
NUSFLVDRGMLEPRKRVALLBELIGLIAAGVFTTPTATPIARVADAFMSMAO  
AOHLGKVLTLTGDEPVQIRIPTHAGAGSSTDRLDLDLASAPARAAALAEPLRTQ  
VSOVTRPEIRVGAELFTRLGMSLMAVEENRNIEMSLKLUKLTSTTFLSTSPALLLA  
ONILDALATLALSERVAENLRAGVQNDVSSGADQDWEIITAL"  
7640..8920  
/gene="epoa"  
/note="Region: beta-ketoacyl synthase/malonyl-CoA  
decarboxylase"  
9236..10201  
/gene="epoa"  
/note="Region: acetate acyltransferase"  
10529..11428  
/gene="epoa"  
/note="Region: enoyl reductase"  
11549..11764  
/gene="epoa"  
/note="Region: acyl carrier protein"  
11872..16104  
/gene="epop"  
11872..16104  
/gene="epop"  
/note="BPOS P; contains condensation/heterocyclization,

cysteine adenylation, putative NAD(P)H oxidase, and  
peptidyl carrier domains"  
/codon\_start=1  
/transl\_table=11  
/product="nonribosomal peptide synthetase"  
/protein\_id="AAF26925.1"  
/db\_xref="GI:6724259"  
/translation="WTINQLNELHEHGIKLAADGERLQIAKNAALNPILARISEH  
KSTILTMRORLPESIVPAPAPERAPPLTDIOESYMLGRTGATVPSGIHAYRED  
CTDLDPRLSAPFRKVAHMDLRAHLPDMQVLEPKVDADIEIDLRGSTRREA  
RLVSLRDSMHSRIYDTERPLYHVAVALDEORQLVSLINDVLSLITFDML  
SEYDEPSTLPLVLESTYRDYLALESRKSEHOSMYMKRIALEPPPTLPKAD  
PSTLKEIFRHEQMLPSDSWGLRKRRGKELPTGVILAAFSVIGRMASPEFL  
NITLENRLPVRVNDITGDFSMVLIDITDRDSFEORAKIOEQLMEAMDCDS  
GIVQREARAVLIGORGALFVYLTSALNOQVGTSLQRLGTPYTSOTPOLDDH  
OLYEHGDVLVAMDIVDGVFPDLDLMEAVFLRLTEPEBOYRSLPAPOLE  
ARASANTNALLSEHTLHGLPAPVBEQLPMOLAVASAKTITYELBSRSRLARLR  
EGKARPTLVAVMEKMEQYVAVALVLESNAVPIADLPAERIHILDHGEKLV  
LTQWLDKLSMPGRIQLLVSEAGVEDGQPPMPITQPSDLVAVIYTGSGTLPK  
GVMI DHRGAVNTIIDLINERFEIGPDRVIALSLSFDISVDVFGILAAQGTIVPDA  
SKLRDPAHMAELIEREKTVNNSPALRMVLEHREGRDPSIARSLRLSGDWIPV  
GLPGLQAIRGVSVISLGATSEASISGVPRVNDLSMSIPIYGRILNOTPHVD  
EALRRPVVPGOLYIGVGLALGYMRDEKTRKSLVHPSTGERLYTKGLKRLPD  
ENIERMGEEDNOIKLRGRVLEGEIEETLKHNPVRDVIVPVGDAKRLDLAVVP  
EGTRRRAEQDASLTERIDAAHAAEDLSDEGRVQFKLARHGLDLGKVPVLD  
TGODPREAGLVYARRSRVTFLEAPIFVEHRELSGLSEVPEHGLATLPFRVPSAG  
STPYQTAAYVAGRIEGVEDGFYYPHERLLKLSHGIERGHAVQNDVDEBA  
PNLIFVGRIDAIESLYGSSREPCLEAGYMAOLMEAPQNGICVGVGFNFOVR  
PVLDRHSDDVYHGMKGRVDPROVCTTQQDSSPRATRGAPRGHQRPAIMLRD  
FLRTKLPEYMPVTFVEIDLALPLTSNGVDRKALKERKDTSPRHSGHAPDLAEEL  
LVAVREVLGLEVGLQQSPVLDLGTSLIHYMRSLLOKRLDRLAILELPOYPLGSS  
LASGLRDSRLDQRPNNQDRVEVRKRRRS"  
16251..21749  
/gene="epob"  
16251..21749  
/gene="epob B"  
/note="EPOS B"  
/codon\_start=1  
/transl\_table=11  
/product="polyketide synthase"  
/protein\_id="AAF26920.1"

Query Match 100.0%; Score 4266; DB 1; Length 68750;  
Beet Local Similarity 100.0%; Pred. No. 0;  
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGGGATTCCTCCATCGAGCGCGGACCCGAAATCCGATTGCGATGTCGAGACGAGT 60  
Db 7610 GTGGGGGATTCCTCCATCGAGCGCGGACCCGAAATCCGATTGCGATGTCGAGACGAGT 7669  
QY 61 TGCCGCTGCGCGGATGCGATCGATCGAGCGGATTCTGGAAGCTCCTCGAGGGCTCG 120  
Db 7670 TGCCGCTGCGCGGATGCGATCGATCGAGCGGATTCTGGAAGCTCCTCGAGGGCTCG 7729  
QY 121 CGCGACACCGTGGGCGAGTCCCGCCGACGCTGGATGCGACGCGTGTGTTGATCC 180  
Db 7730 CGCGACACCGTGGGCGAGTCCCGCCGACGCTGGATGCGACGCGTGTGTTGATCC 7789  
QY 181 GACCCCGATGCCCGGGGAAAGCGCCCGTTACGCGCGCATCTTTCTTGAAGCAGTACGCC 240  
Db 7790 GACCCCGATGCCCGGGGAAAGCGCCCGTTACGCGCGCATCTTTCTTGAAGCAGTACGCC 7849  
QY 241 TGCTTGAAGCGCTCTTCTTGGGCGATCTCGCTCGCGAAGCGCTGAGTGAAGCCCTGCA 300  
Db 7850 TGCTTGAAGCGCTCTTCTTGGGCGATCTCGCTCGCGAAGCGCTGAGTGAAGCCCTGCA 7909  
QY 301 CATCGACTCTTGCTGAGAGTGCTGAGAGCGGCTGGAAGACGCCGCGCATGCTTCATCG 360  
Db 7910 CATCGACTCTTGCTGAGAGTGCTGAGAGCGGCTGGAAGACGCCGCGCATGCTTCATCG 7969  
QY 361 GCGCTCGTCGTPACGGAACGCGAGTCTTTCATCGGGATTCGCGCCCTCCGAATATAGAGCC 420  
Db 7970 GCGCTCGTCGTPACGGAACGCGAGTCTTTCATCGGGATTCGCGCCCTCCGAATATAGAGCC 8029

QY 421 GCGCTGCCGCAAGCAGCGCGCTCCGACAGATTCAGCGCTCAGCGCGCTGGGACGATG 480  
Db 8030 GCGCTGCCGCAAGCAGCGCGCTCCGACAGATTCAGCGCTCAGCGCGCTGGGACGATG 8089  
QY 481 CCCAGCTTCGAGAGCGGCGCCGAATCTTGATTCCTCCGAGCTGCGAGGCGCTGTGTCGG 540  
Db 8090 CCCAGCTTCGAGAGCGGCGCCGAATCTTGATTCCTCCGAGCTGCGAGGCGCTGTGTCGG 8149  
QY 541 GTGATACGGCTTATTCGCTCCGCTGGTGGCCGTTACCTGGCCTGACAGAGCTTGGCC 600  
Db 8150 GTGATACGGCTTATTCGCTCCGCTGGTGGCCGTTACCTGGCCTGACAGAGCTTGGCC 8209  
QY 601 TCCGGGATGCTCCACGCGCCCTGCTGCTGGTGGGATTCGATGTTGTCGCCGACACC 660  
Db 8210 TCCGGGATGCTCCACGCGCCCTGCTGCTGGTGGGATTCGATGTTGTCGCCGACACC 8269  
QY 661 CTGCTGTGCTCTCGAAGACCCCGGCGCTGGCCAGGACGCTCCGTGCAAGGCATTTTCG 720  
Db 8270 CTGCTGTGCTCTCGAAGACCCCGGCGCTGGCCAGGACGCTCCGTGCAAGGCATTTTCG 8329  
QY 721 GCGAGAGCGCATGAGGTTTCGACGAGGCGGAGGCGCGCGCTGCTGCAAGCGGCTC 780  
Db 8330 GCGAGAGCGCATGAGGTTTCGACGAGGCGGAGGCGCGCGCTGCTGCAAGCGGCTC 8389  
QY 781 AGTGAAGCCCGCGCGGACGCGGATCGGATATTGGCGGTGATTCAGAGATCCCGCATCAAT 840  
Db 8390 AGTGAAGCCCGCGCGGACGCGGATCGGATATTGGCGGTGATTCAGAGATCCCGCATCAAT 8449  
QY 841 CACGACGGTGCAGACAGCGGCTTCTGACCGTGCAGAACGAGAGCTCCCAAGAAATGCTGCTG 900  
Db 8450 CACGACGGTGCAGACAGCGGCTTCTGACCGTGCAGAACGAGAGCTCCCAAGAAATGCTGCTG 8509  
QY 901 AAACGGCGCTGGGCGGACGCGGCTGCGCGCGCTCTTCGAGGATTAATGAGGCAAC 960  
Db 8510 AAACGGCGCTGGGCGGACGCGGCTGCGCGCGCTCTTCGAGGATTAATGAGGCAAC 8569  
QY 961 GGCACGGGACGACGCTTGTGTAACCCCATCGAAATCCAGCTCGAATGCGGTATACGGC 1020  
Db 8570 GGCACGGGACGACGCTTGTGTAACCCCATCGAATCCAGCTCGAATGCGGTATACGGC 8629  
QY 1021 CTGCGGCGAGATGTCGCCACGCGCTGCTGATCGGGTTCGTGAAGACCACTTGGCCAT 1080  
Db 8630 CTGCGGCGAGATGTCGCCACGCGCTGCTGATCGGGTTCGTGAAGACCACTTGGCCAT 8689  
QY 1081 CCTGAGTATGCTGGGGATTCATCGGCTGCTGAAGCTGCTTGCCTTACACGCGG 1140  
Db 8690 CCTGAGTATGCTGGGGATTCATCGGCTGCTGAAGCTGCTTGCCTTACACGCGG 8749  
QY 1141 CAGATTCTTGCGGACCTTCACGCGCGAGCGCTGAACCCCGGATCTCATGAGGATGATCTT 1200  
Db 8750 CAGATTCTTGCGGACCTTCACGCGCGAGCGCGCTGAACCCCGGATCTCATGAGGATGATCTT 8809  
QY 1201 CGGCTGACCGTCAAGCGCGCCCGGACACCGTGGCCGACCTGGAATACGCGCGACGCGG 1260  
Db 8810 CGGCTGACCGTCAAGCGCGCCCGGACACCGTGGCCGACCTGGAATACGCGCGACGCGG 8869  
QY 1261 GGGGTGAGCTGTTCTGGGATGAGCGGGAACCAAGCCGACGCTGTGCTGGAAGAGCGCGG 1320  
Db 8870 GGGGTGAGCTGTTCTGGGATGAGCGGGAACCAAGCCGACGCTGTGCTGGAAGAGCGCGG 8929  
QY 1321 GCGCGAGAGTGCACACGCGCGCGCGGAGGACCGGACAGAGCTGTGCTGTGCGCA 1380  
Db 8930 GCGCGAGAGTGCACACGCGCGCGCGGAGGACCGGACAGAGCTGTGCTGTGCGCA 8989  
QY 1381 AGGACCGGCTGACCGCTTGATGACACAGGCGGCGCGCTGCGGACCATTTGAGACCTTAC 1440  
Db 8990 AGGACCGGCTGACCGCTTGATGACACAGGCGGCGCGCTGCGGACCATTTGAGACCTTAC 9049  
QY 1441 CCTTGGAGAGTCTGGGCGAGATGTCGCTTCACTGGCGACAGAGCGGACGCGCATGAGAG 1500  
Db 9050 CCTTGGAGAGTCTGGGCGAGATGTCGCTTCACTGGCGACAGAGCGGACGCGCATGAGAG 9109



Db 11270 CCGGCGGGGCTCCGCTCTTGGAGAGAGCTCTCGGCTGATCGCGGAGGGCTGTT 11329  
Qy 3721 ACCCTCCCCCATCGAGAGCTCCCGATCGCCGCTGCGCGAGTGGCTTCGGAGCATG 3780  
Db 11330 ACCCTCCCCCATCGAGAGCTCCCGATCGCCGCTGCGCGAGTGGCTTCGGAGCATG 11389  
Qy 3781 GCGGAGCGGAGCATCTTGGAGAGCTCTGATCTCAAGCTGGGTGACCCGAGGTCAGATC 3840  
Db 11390 GCGGAGCGGAGCATCTTGGAGAGCTCTGATCTCAAGCTGGGTGACCCGAGGTCAGATC 11449  
Qy 3841 GGTATTCACACCCAGAGAGCGCGCGCTGTCACCGGGGATCGGAGCTCTGACAGG 3900  
Db 11450 GGTATTCACACCCAGAGAGCGCGCGCTGTCACCGGGGATCGGAGCTCTGACAGG 11509  
Qy 3901 CTGCGAGCTGAGCG 3960  
Db 11510 CTGCGAGCTGAGCG 11569  
Qy 3961 GTCCTGCGAGGCTGCTGCGAGCGCGCGAGATCAAGGTCGGCGCGAGCGCTGTCACCGCG 4020  
Db 11570 GTCCTGCGAGGCTGCTGCGAGCGCGCGAGATCAAGGTCGGCGCGAGCGCTGTCACCGCG 11629  
Qy 4021 CTGCGAGCTGAGCTGCTGATGAGCGCGTGAAGCTGCGCATCTGATTCAGAGCGAGCTCAAG 4080  
Db 11630 CTGCGAGCTGAGCTGCTGATGAGCGCGTGAAGCTGCGCATCTGATTCAGAGCGAGCTCAAG 11689  
Qy 4081 CTGAGGCTGTCAGAGAGCTTCTGTCAGAGTCCCGCATATCGCTTGGCGCGCAAG 4140  
Db 11690 CTGAGGCTGTCAGAGAGCTTCTGTCAGAGTCCCGCATATCGCTTGGCGCGCAAG 11749  
Qy 4141 CTGTTGATGCTCTGCGCAGAGCTCTCTCTTGGAGCGGCTGCGCGCGAGAGACTCAAG 4200  
Db 11750 CTGTTGATGCTCTGCGCAGAGCTCTCTCTTGGAGCGGCTGCGCGCGAGAGACTCAAG 11809  
Qy 4201 GCGAGCGCTGCAAAAGAGCTTGTCTCATCGCGCGCGAGATCAAGACTGGGAAATCATGCGC 4260  
Db 11810 GCGAGCGCTGCAAAAGAGCTTGTCTCATCGCGCGCGAGATCAAGACTGGGAAATCATGCGC 11869  
Qy 4261 CTATGA 4266  
Db 11870 CTATGA 11875

RESULT 2  
ARI93029 68750 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 1 from patent US 6346404.  
ACCESSION ARI93029  
VERSION ARI93029.1 GI:20238994  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)  
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and  
Goriach,J.  
TITLE Genes for the biosynthesis of epoethiones  
JOURNAL Patent: US 6346404 A 1 12-FEB-2002;  
FEATURES  
source  
1. 68750  
/organism="unknown"  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN

Query Match 100.0%; Score 4266; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGAGATGCTCCATGAGAGCGCGAGCCGAGAGATCGATTGCGATGTCGAGCGGAGT 60  
Db 7610 GTGGGAGATGCTCCATGAGAGCGCGAGCCGAGAGATCGATTGCGATGTCGAGCGGAGT 7669  
Qy 61 TGCCGCTGCGCGGCTGCGGATCGATCGATCGAGCGGCTTCTGAGACGCTCTCGAGGCGTCG 120

Db 7670 TGCCGCTGCGCGGCTGCGGATCGATCGATGAGCGGCTTCTGAGACGCTCTCGAGGCGTCG 7729  
Qy 121 GCGAGACCGTCCGGGCGAGTCCCGCGAGAGCGTGGATGAGAGGCGTGGATTGATCCC 180  
Db 7730 GCGAGACCGTCCGGGCGAGTCCCGCGAGAGCGTGGATGAGAGGCGTGGATTGATCCC 7789  
Qy 181 GACCCGATGCGCGGAGAGAGCGCCGTTACGCGCGCATCTTCTCGAGGAGTGAAGCC 240  
Db 7790 GACCCGATGCGCGGAGAGAGCGCCGTTACGCGCGCATCTTCTCGAGGAGTGAAGCC 7849  
Qy 241 TGCTTGACAGCTCTCTTCTTCTGCGATCTCGCTCGAGAGCGCTGCGGATGAGACCTGCA 300  
Db 7850 TGCTTGACAGCTCTCTTCTTCTGCGATCTCGCTCGAGAGCGCTGCGGATGAGACCTGCA 7909  
Qy 301 CATGAGCTTCTGAGAGGTGTGTGAGAGCGCTGAGAGAGCGCGGATGAGCTCAATCG 360  
Db 7910 CATGAGCTTCTGAGAGGTGTGTGAGAGCGCTGAGAGAGCGCGGATGAGCTCAATCG 7969  
Qy 361 GCGCTGCTGAGTACGAGAAAGGAGTGTTCATCGGATCGGCGCGTCCGAAATATGAGGCC 420  
Db 7970 GCGCTGCTGAGTACGAGAAAGGAGTGTTCATCGGATCGGCGCGTCCGAAATATGAGGCC 8029  
Qy 421 GCGCTGCTGAGTACGAGAGCGGCTCCGAGAGATCGACGCTCATGCGCGGCTGCGGAGCATG 480  
Db 8030 GCGCTGCTGAGTACGAGAGCGGCTCCGAGAGATCGACGCTCATGCGCGGCTGCGGAGCATG 8089  
Qy 481 CCGAGCGTCCGAGAGCGGCGCGAATCTGATGAGCGCTCGGAGCGAGAGCGGCTGAGCGG 540  
Db 8090 CCGAGCGTCCGAGAGCGGCGCGAATCTGATGAGCGCTCGGAGCGAGAGCGGCTGAGCGG 8149  
Qy 541 GTGATATCGGCTATGCTGCTCTGCTGAGCGCTTCAATCGAGCTGAGCTGAGCGG 600  
Db 8150 GTGATATCGGCTATGCTGCTCTGAGCGCTTCAATCGAGCTGAGCTGAGCGG 8209  
Qy 601 TCCGAGAAATGCTCCAGCGCGCTGCTGAGAGGATGCTGATGTTGTCGCGCAAGC 660  
Db 8210 TCCGAGAAATGCTCCAGCGCGCTGCTGAGAGGATGCTGATGTTGTCGCGCAAGC 8269  
Qy 661 CTGCTGCTGCTGCTGAGAGAGCGCGCGCGCTGCGAGAGAGCGTCCGAGAGCAATTTTCG 720  
Db 8270 CTGCTGCTGCTGCTGAGAGAGCGCGCGCGCTGCGAGAGAGCGTCCGAGAGCAATTTTCG 8329  
Qy 721 GCGAGGCGGATGAGGCTTCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
Db 8330 GCGAGGCGGATGAGGCTTCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8389  
Qy 781 AGTGAAGCGCGCGGAGAGCGGATCGAGATATGCGGATGATTCAGAGATCCGCAATCAAT 840  
Db 8390 AGTGAAGCGCGCGGAGAGCGGATCGAGATATGCGGATGATTCAGAGATCCGCAATCAAT 8449  
Qy 841 CACGAGCGTGGAGACACCGGCTTGAACCGGCGGAGAGCGGAGAGCGGAGAGAGAGAGAGAG 900  
Db 8450 CACGAGCGTGGAGACACCGGCTTGAACCGGCGGAGAGCGGAGAGCGGAGAGAGAGAGAGAG 9509  
Qy 901 AAAGGAGCGCTGAG 960  
Db 8510 AAAGGAGCGCTGAG 8569  
Qy 961 GGCAGGAG 1020  
Db 8570 GGCAGGAG 8629  
Qy 1021 CTGCGGAGAGATGTCGAG 1080  
Db 8630 CTGCGGAGAGATGTCGAG 8689  
Qy 1081 CCTGATGAGCTGAGAGAGATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
Db 8690 CCTGATGAGCTGAGAGAGATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8749  
Qy 1141 CAGATTCCTGAGAGCTTCAAG 1200

Dp	8750	AAGATTCTCGGACA	CTTC	CAAGCGGAGCGCTGAACCCCGGGATCTCATGGGGTGATCTTT	8809
Qy	1201	CGGCTGAC	CTTC	ACGCGCGCCCGGACACCTGTGGCCGGACTGGAAATAGCCCGCGACGGGCG	1260
Dp	8810	CGGGCTGAC	CTTC	ACGCGCGCGCCCGGACACCTGTGGCCGGACTGGAAATAGCCCGCGACGGGCG	8869
Qy	1261	GGGGTGAG	CTG	CTTCCGAGTGAAGGGGGGACCAACCGGACCTGTGGGTCTGGAAAGAGCGCG	1320
Dp	8870	GGGGTGAG	CTG	CTTCCGAGTGAAGGGGGGACCAACCGGACCTGTGGGTCTGGAAAGAGCGCG	8929
Qy	1321	CGCGCGAC	AGT	GCACACCGCGCGCGCGGAGCGACCGGACAGAGCTGTGGTCTTCCGCA	1380
Dp	8930	CGCGCGAC	AGT	GCACACCGCGCGCGCGGAGCGACCGGACAGAGCTGTGGTCTTCCGCA	8989
Qy	1381	AGGACCCG	CGT	CAGCCCTTGGATGCAACGCGGCGCGGCTGGCGACATCTGAGAGCTTAC	1440
Dp	8990	AGGACCCG	CGT	CAGCCCTTGGATGCAACGCGGCGCGGCTGGCGACATCTGAGAGCTTAC	9049
Qy	1441	CCTTCGAG	AGT	CTTCCGAGTGAAGGGGGGACCAACCGGACCTGTGGGTCTGGAAAGAGCGCG	1500
Dp	9050	CCTTCGAG	AGT	CTTCCGAGTGAAGGGGGGACCAACCGGACCTGTGGGTCTGGAAAGAGCGCG	9109
Qy	1501	CACCGGCT	CGG	GGGTGGCGGACGTCGAGGAGGGGGCTGGCGGACGCCCTTGAAGCTTACG	1560
Dp	9110	CACCGGCT	CGG	GGGTGGCGGACGTCGAGGAGGGGGCTGGCGGACGCCCTTGAAGCTTACG	9169
Qy	1561	GGCGAGGAG	ACAG	CGTCCGCCGGTGGCGAGATATCGCCATTCTCACCGCGGCAAG	1620
Dp	9170	GGCGAGGAG	ACAG	CGTCCGCCGGTGGCGAGATATCGCCATTCTCACCGCGGCAAG	9229
Qy	1621	CTCCGCTTT	CT	CTTCAACCGGACAGGGGGGCGGACACCTGGGACATGGGCGGTGGCTGAC	1680
Dp	9230	CTCCGCTTT	CT	CTTCAACCGGACAGGGGGGCGGACACCTGGGACATGGGCGGTGGCTGAC	9289
Qy	1681	GATGATAT	GGT	CCGCGGCTTCCGCGAGGCGCTTTCGACCTGTGCGAGGACTTTTCAACGAGAG	1740
Dp	9280	GATGATAT	GGT	CCGCGGCTTCCGCGAGGCGCTTTCGACCTGTGCGAGGACTTTTCAACGAGAG	9349
Qy	1741	CTCGACCG	GGG	CGCTCCGCGAGGTATGTGGGCGGACCGGACAGCGTGCAGCGCGCTG	1800
Dp	9350	CTCGACCG	GGG	CGCTCCGCGAGGTATGTGGGCGGACCGGACAGCGTGCAGCGCGCTG	9409
Qy	1801	CTCGAC	CCAG	ACAGCCTTACCCAGCGCGCGCTGTTACCTTCCGAATATGCGCTGGCGCG	1860
Dp	9410	CTCGAC	CCAG	ACAGCCTTACCCAGCGCGCGCTGTTACCTTCCGAATATGCGCTGGCGCG	9469
Qy	1861	CTGTGGCG	AGT	GTGAGGCGGAGCTTGGTCCGCGGACATAGCATGGGTGAGCTG	1920
Dp	9470	CTGTGGCG	AGT	GTGAGGCGGAGCTTGGTCCGCGGACATAGCATGGGTGAGCTG	9529
Qy	1921	GTGGCTG	CGT	TCGCGGAGCGGTGTTCTCGCTTGAAGACGCGGTGTTCTTGTGGACTGCG	1980
Dp	9530	GTGGCTG	CGT	TCGCGGAGCGGTGTTCTCGCTTGAAGACGCGGTGTTCTTGTGGACTGCG	9589
Qy	1981	CGCGGGCG	CGT	GCAGGCGTGCAGGCGCTGCGCGGCGGAGTGGTGCATGAGGCGCG	2040
Dp	9590	CGCGGGCG	CGT	GCAGGCGTGCAGGCGCTGCGCGGCGGAGTGGTGCATGAGGCGCG	9649
Qy	2041	GAGGCGCA	TG	GGGCTGCGGTTGGGCGCGCGACGACGAGGTGGGTGCGATGCGCGCGCT	2100
Dp	9650	GAGGCGCA	TG	GGGCTGCGGTTGGGCGCGCGACGAGGTGGGTGCGATGCGCGCGCT	9709
Qy	2101	AACGCT	CCGGA	CAAGGTGATCGCGGCGCGGCAACCCGTCAGTACGATGCGCGCG	2160
Dp	9710	AACGCT	CCGGA	CAAGGTGATCGCGGCGCGGCAACCCGTCAGTACGATGCGCGCG	9769
Qy	2161	GCAGT	AGC	CGCGCGGAGGCGCGGACCAAGCGCTCCAGTCTTCGATGCTTCCACTCA	2220
Dp	9770	GCAGT	AGC	CGCGCGGAGGCGCGGACCAAGCGCTCCAGTCTTCGATGCTTCCACTCA	9829
Qy	2221	CCGCTCA	TG	CGCCGATGCGCGGCGGAGCGGAGCGGCTTCCGCGGCGTGTGGCGAGTACGCTACCGG	2280
Dp	9830	CCGCTCA	TG	CGCCGATGCGCGGCGGAGCGGAGCGGCTTCCGCGGCGTGTGGCGAGTACGCTACCGG	9889

QY	2281	CGGCGCGCATCGTCCCGGTGAGCAATCTGAGCGGGAAGCTTGGCAACAAGAGTGAC	2340
Db	9990	CGGCGCGCATCGTCCCGGTGAGCAATCTGAGCGGGAAGCTTGGCAACAAGAGTGAC	9949
QY	2341	TCGCGCGGACATTGGGTGCGCCACGCGCGAGAGGTGTGCGCTTCGCGATGAGTGAAG	2400
Db	9950	TCGCGCGGACATTGGGTGCGCCACGCGCGAGAGGTGTGCGCTTCGCGATGAGTGAAG	10009
QY	2401	GCGCTGACCGCGGCGCGGTGCGGACACTTCCTGAGAGTGGTCCGAAATCGACGCTGCTC	2460
Db	10010	GCGCTGACCGCGGCGCGGTGCGGACACTTCCTGAGAGTGGTCCGAAATCGACGCTGCTC	10069
QY	2461	GCGCTGGTGTGCTCTGTCGATATGCGGAGCGGCCGCGCGCTCTGCGATCTGTCGCGCT	2520
Db	10070	GCGCTGGTGTGCTCTGTCGATATGCGGAGCGGCCGCGCGCTCTGCGATCTGTCGCGCT	10129
QY	2521	GGGCGTGAAGAGCCGCGGACCGTGTCTCGAGAGCGCTCGCGGCTCTGGGCGCTCGTGGC	2580
Db	10130	GGGCGTGAAGAGCCGCGGACCGTGTCTCGAGAGCGCTCGCGGCGCTCTGGGCGCTCGTGGC	10189
QY	2581	CTGGTCTCCGAGGCGCGGCTCTTCCCTCAGGGGGGGCGCGGGGTGCGCTGGCCACGTCAC	2640
Db	10190	CTGGTCTCTGAGGCGCGGCTCTTCCCTCAGGGGGGGCGCGGGGTGCGCTGGCCACGTCAC	10249
QY	2641	CCTTGGCAGCGCGAGCGCTTACTGGATCGACAGAAAGCCGACGACGCGCGCGTGGCGAC	2700
Db	10250	CCTTGGCAGCGCGAGCGCTTACTGGATCGACAGAAAGCCGACGACGCGCGCGTGGCGAC	10309
QY	2701	CGCGGTCTCCGGGAGCGGGTCAAGAGAGTGTGAGAGAGGGGGGGCGGTGCGCGCGCGC	2760
Db	10310	CGCGGTCTCCGGGAGGGGTCAAGAGAGTGTGAGAGAGGGGGGGCGGTGCGCGCGCGC	10369
QY	2761	GACCGGCGCAGCGCTCGGCTCGACCAATCCGCGCGCCGAGAGCGGACCGCGGAGAGAGTCT	2820
Db	10370	GACCGGCGCAGCGCTCGGCTCGACCAATCCGCGCGCCGAGAGCGGACCGCGGAGAGAGTCT	10429
QY	2821	GAGCGCGCGCGCGACCGTCCGTTCCGGCTTCGAGATCGATGAGCCAGGCGTCTTGATCAC	2880
Db	10430	GAGCGCGCGCGCGACCGTCCGTTCCGCGCTCGAGATCGATGAGCCAGGCGTCTTGATCAC	10489
QY	2881	CTCGTGCTTGGGTTCACGAGAGCGCGCGCCGCTGTGGAGAGAGTCCAGATTCGCGCTC	2940
Db	10490	CTCGTGCTTGGGTTCACGAGAGCGCGCGCCGCTGTGTGGAGAGAGTCCAGATTCGCGCTC	10549
QY	2941	GACGCGCGGGGCTCAGACTTCAATGATGTCCAGCTTCGCGCTGCGATGCTGCCACGAC	3000
Db	10550	GACGCGCGGGGCTCAGACTTCAATGATGTCCAGCTTCGCGCTGCGATGCTGCCACGAC	10609
QY	3001	CTGCGCGGGAAGGCCCAACCTCCGCGCTGCTCGAGAGGAGAGTGGCGCGGGGCGAATCGTC	3060
Db	10610	CTGCGCGGGAAGGCCCAACCTCCGCGCTGCTCGAGAGGAGAGTGGCGCGGGGCGAATCGTC	10669
QY	3061	GCGGTGGCGAGGAGCGTGAACCGGCTCTGTGTGTGGGCGCAACCGGCTATCGGCTTCGCGC	3120
Db	10670	GCGGTGGCGAGGAGCGTGAACCGGCTCTGTGTGTGGGCGCAACCGGCTATCGGCTTCGCGC	10729
QY	3121	GAGCGCTTGTCTACCCACGTCACACAGTGGGCTGCGCTGTGTGCTCTGGGCTCAAGGCG	3180
Db	10730	GAGCGCTTGTCTACCCACGTCACACAGTGGGCTGCGCTGTGTGCTCTGGGCTCAAGGCG	10789
QY	3181	CTCTCGGCGATTCAGAGGCGGCGCGCATGCGCGTGTGCTACTGACGGGCAATGTACCGGCTC	3240
Db	10790	CTCTCGGCGATTCAGAGGCGGCGCGCATGCGCGTGTGCTACTGACGGGCAATGTACCGGCTC	10849
QY	3241	GACGAAATAGCCCGCTTCAGCCCGGGGAGCGGGGTCTGATTCATTCGCGGACCGCGGCG	3300
Db	10850	GACGAAATAGCCCGCTTCAGCCCGGGGAGCGGGGTCTGATTCATTCGCGGACCGCGGCG	10909
QY	3301	GTCGCTTCCGCGGTGAGAGTGGGGCGCAGCATGTGGAGCGGAGTTCATGAGAGCGGC	3360
Db	10910	GTCGCTTCCGCGGTGAGAGTGGGGCGCAGCATGTGGAGCGGAGTTCATGAGAGCGGC	10969

QY	3361	GGCAGCGCCGAGAAAGCGGCTACCTGGAAGTCGTGGGCGCTGGGGATGATGAGCCATTCC	3420
Db	10970	GGCAGCGCCGAGAAAGCGGCTTACTTGAAGTCGTGGGCGCTGGGGATGATGAGCCATTCC	110229
QY	3421	CGCTCGGACCGGTTTCGTGCGCGAGCTGCGCGCGTGAACGCGGCGGAGGAGTAGACGTC	3480
Db	11030	CGCTCGGACCGGTTTCGTGCGCGAGCTGCGCGCGTGAACGCGGCGGAGGAGTAGACGTC	11089
QY	3481	GTGCTCAACTCGCTCTGGGCGAGCTATGACAGAATTTCATCTCTGCGATGCGAC	3540
Db	11090	GTGCTCAACTCGCTCTGGGCGAGCTATGACAGAAGTTTCATCTCTGCGATGCGAC	11149
QY	3541	GGCGGGTTTGTGAGAGCTCGGCAAGCGGACGTGTTACGCGGATPAACAGAGCTCGGCTGCGG	3600
Db	11150	GGCGGGTTTGTGAGAGCTCGGCAAGCGGACGTGTTACGCGGATPAACAGAGCTCGGCTGCGG	11208
QY	3601	CCGTTCTGTGCGCAATCTCTCTTCTCGCTGTGGATCTCGGGGGATGATGCTCGACCG	3660
Db	11210	CCGTTCTGTGCGCAATCTCTCTTCTCGCTGTGGATCTCGGGGGATGATGCTCGACCG	11266
QY	3661	CGGCGCGGGTCCGTGCGCTTTGGAGAGCTCCTCGGCTGATTCGCGCGAGGCGTGTTC	3720
Db	11270	CGGCGCGGGTCCGTGCGCTTTGGAGAGACTCCTCGGCTGATTCGCGCGAGGCGTGTTC	11328
QY	3721	ACCCCTCCCCCAATCGGAGAGCGCTCCGGAATGCCCTGTGCCCGATCGTTCCGGAACAG	3780
Db	11330	ACCCCTCCCCCAATCGGAGAGCGCTCCGATGCCCTGTGCCCGATCGTTCCGGAACAG	11388
QY	3781	GCGCAGCGCGCAGCATCTTGGGAAAGCTGTACTACGCTGGGGTACCCGAGAGTCCAGATC	3840
Db	11390	GCGCAGCGCGCAGCATCTTGGGAAAGCTGTACTACGCTGGGGTACCCGAGAGTCCAGATC	11448
QY	3841	CGTATTCACACCAACGAGGCGCGGCTGTCAACGCGGGATCGGAACTGTGTCGACAG	3900
Db	11450	CGTATTCACACCAACGAGGCGCGGCTGTCAACGCGGGATCGGAACTGTGTCGACAG	11508
QY	3901	CTCGCGTCAGCTGGGCGCGGCGCGCGCGCGCGCGCGCTGTGAAGCGTTCTCCGCTAGCAG	3960
Db	11510	CTCGCGTCAGCTGGGCGCGGCGCGCGCGCGCGCGCGCTGTGAAGCGTTCTCCGCTAGCAG	11568
QY	3961	GTCCTCGCAGGTGCTGCGCAGCGCCGGAATCAAGTTCGCGCGGAGCGCTGTCAACCGC	4020
Db	11570	GTCCTCGCAGGTGCTGCGCAGCGCCGGAATCAAGTTCGCGCGGAGCGCTGTCAACCGC	11628
QY	4021	CTCGGCAATGACCTGCTCATAGCCGTGAGCTGCGCAATCGTATCGAGCGAGCTTCAG	4080
Db	11630	CTCGGCAATGAGCTGCTCATAGCCGTGAGCTGCGCAATCGTATCGAGCGAGCTTCAG	11688
QY	4081	CTGAAGCTGTTCGAGAGAGCTTCTGTCCAAGTCCCCCAATATGCGCTTGTGGCCCAAAAC	4140
Db	11690	CTGAAGCTGTTCGAGAGAGCTTCTGTCCAAGTCCCCCAATATGCGCTTGTGGCCCAAAAC	11749
QY	4141	CTGTTGGATGCTTCGCGCAACAGCTCTCTCTGTGAGCGGGTGGCGCGGAGAACTTACGG	4200
Db	11750	CTGTTGGATGCTTCGCGCAACAGCTCTCTCTGTGAGCGGGTGGCGCGGAGAACTTACGG	11808
QY	4201	GCAGGCGTGCAAAACGACTTCTGTCTCATTCGCGCGCAGATCAAGACTTGGAAATCATTTGCC	4260
Db	11810	GCAGGCGTGCAAAACGACTTCTGTCTCATTCGCGCGCAGATCAAGACTTGGAAATCATTTGCC	11868
QY	4261	CTATGCA 4266	
Db	11870	CTATGCA 11875	

RESULT	3
LOCUS	AR199551
DEFINITION	Sequence 68750 bp DNA linear PAT 20-APR-2002
ACCESSION	AR199551
VERSION	AR199551.1 GI:20249625
KEYWORDS	.
SOURCE	Unknown.

ORGANISM	unknown.
REFERENCE	unclassified.
AUTHORS	1 (bases 1 to 68750)
TITLE	Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Gorlach,J.
JOURNAL	Genes for the biosynthesis of epochnones
FEATURES	Patent: US 635457-A 1 12-MAR-2002;
source	Location/Qualifiers
	1..68750
BASE COUNT	/organism="unknown" 9596 a 22456 c 25539 g 11159 t
ORIGIN	

Query Match	100.0%;	Score 4266;	DB 6;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4266;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 1 GTGGCGATCGTCCCATCGAGCGCGCAGCCGAGATCCGATTGCCGATCTCGAGCGAGT 60

Dh 7610 GTGGCGATCGTCCCATCGAGCGCGCAGCCGAGATCCGATTGCCATCGTCCGAGCGACT 766

130

\_\_\_\_\_

100

121 CGGACACCGTCGGGAGTCCCGCGAACGCTGGATGACGACGCTGGTTCGATCC 180

Db 7730 CGCGACACCGTCGGGCGAGTCCCCCGCGAACGCTGGGATGCAGCAGCGTGGTTGATCCC 778

181 GACCCGATGCCCCGGGGAAGACGCCCGTTACGCGGCATCTTTCCTGAGCGACGTAGCC 240

Db 7790 GACCCGATGCCCGGGGAGACGCCCGTTACGGCGCATCTTCTGAGCGACGTAGCC 784

241 TGCCTTTCGACGCTCTTCTCTTCTTCGCGCATCTTCGCGCGAAGCGCTGCGGATGGACCGCTGCA 300

2000

---

27

DB 7910 CATCGACTCTTCGAGGAGGCTGGAGGCGCTGGAGAACCGCCGATCGCTCCATCG /96

361 GCGCTCGTGGTACGGAACGGGAGTGTTCATCGGGATCGGCCCGTCCGAATATGAGGCC 420

Db 7970 GCGCTCGTACGGAACGGAGTGTTCATCGGATCGGCCGTCGGAATAGGCC 802

421 GCGCTGCCGCAAGCGACGGCTCCGCAGAGATCGACGCTCATGGCGGCTGGGACGATG 480

8030 GCGCTGCGCAAGCGACGCGTCCGACGATCGACGCTCATGGCGGGCTGGGACGATG 808

[illegible][illegible]

541 GTGATACGGCCCTATTGCTCCCTCGCTGGTGGCCGTCATCTGGCCCTGTACAGAGCTTGGCC 600

Db 8150 GTGGATACGGCCTATTCTGTCCTCGCTGGCCGTTCACTGGCCTGTCAAGCTTGGCC 820

601 TCCGGGATGCTCCACGGCCCTGGCTGGTGGGTATCGCTGATGTTGTCCCGACACC 660

Dh 8210 TCCGGGATGCTCCACGGCCCTGGCTGGTGGGTATCCGCTGATGTGTGCCCGAGCACC 826

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

Db 8330 GCGAGGCCGATGGGTTCCGACGAGCGGAGGGTGCCGCCGTCCGTGCTCAAGCCGCTC 8338

781 AGTGAGCCCGCGGACGGCATCGGATATGGCGGTGATTCGAGGATCCGGCATCAAT 840

Db 8390 AGTGAGCCCGCGGACGGCATCGGATATTGGCGTGATTCGAGGATCCGGATCAAT 844



OY	841	AACGACGGGTGGACACAGCGGCTTCAACCGTGGCGGAAACGGGAACTCCCAAGAAATGTCCTG	900
Db	8450	CACGACGGGTGGACACAGCGGCTTCAACCGTGGCGGAAACGGGAACTCCCAAGAAATGTCCTG	8509
OY	901	AAACGGGGCCCTGGGGGAGACGACAGGCTGCGCCCGCTTCTTCGTGGGGTTATGTCGAGGACAC	960
Db	8510	AAACGGGGCCCTGGGGGAGACGACAGGCTGCGCCCGCTTCTTCGTGGGGTTATGTCGAGGACAC	8569
OY	961	GGCACGGGACGACGCGCTTGTGACCCCATCGAAATCCAAAGCTCTGAATGCGGTATACGGC	1020
Db	8570	GGCACGGGACGACGCGCTTGTGACCCCATCGAAATCCAAAGCTCTGAATGCGGTATACGGC	8629
OY	1021	CTCGGGCGAGATGTCGCCACCGCCGCTGTATCGGTGGGTGAAGACCAACTTGGGCAT	1080
Db	8630	CTCGGGCGAGATGTCGCCACCGCCGCTGTATCGGTGGGTGAAGACCAACTTGGGCAT	8689
OY	1081	CCTGAGTATGCGTGGGGATCATCTGAGGCTGTGTAAGGTCGTTGTCCCTTCAGACGGG	1140
Db	8680	CCTGAGTATGCGTGGGGATCATCTGAGGCTGTGTAAGGTCGTTGTCCCTTCAGACGGG	8749
OY	1141	CAGATTCTCTGGGCACTTCACGCGCAGGCGGCTGAAACCCCGGATCTCAATGGGGTATCTT	1200
Db	8750	CAGATTCTCTGGGCACTTCACGCGCAGGCGGCTGAAACCCCGGATCTCAATGGGGTATCTT	8809
OY	1201	CGGCTGACCTGTCAACGCGCGCCCGGACACACCTGTGGCCGACATGGAATAACGCCGACGGGCG	1260
Db	8810	CGGCTGACCTGTCAACGCGCGCCCGGACACACCTGTGGCCGACATGGAATAACGCCGACGGGCG	8869
OY	1261	GGGGTGAAGCTGTTCCGAGCATGAGCGGAGACAAACGGGCACTGTGATGCTGGAAGAGACGGCG	1320
Db	8870	GGGGTGAAGCTGTTCCGAGCATGAGCGGAGACAAACGGGCACTGTGATGCTGGAAGAGACGGCG	8929
OY	1321	GCGGCGACGTGCAACAACGCGCGGCGCCGAGCGACACCGGACAGACTTGTGGTCTGTGCGCA	1380
Db	8930	GCGGCGACGTGCAACAACGCGCGGCGCCGAGCGACACCGGACAGACTTGTGGTCTGTGCGCA	8989
OY	1381	AGGACCCGCGGTCAAGCCCTGGAATGCAACGAGCGGCGGCTGCGCGACCAATCTTGGAACCTTAC	1440
Db	8990	AGGACCCGCGGTCAAGCCCTGGAATGCAACGAGCGGCGGCTGCGCGACCAATCTTGGAACCTTAC	9049
OY	1441	CCTTCGACGATGTCGGGGGAGATGGGCGTTCAAGTCGTCGGAGACGACGCGACGGAGATGGAG	1500
Db	9050	CCTTCGACGATGTCGGGGGAGATGGGCGTTCAAGTCGTCGGAGACGACGCGACGGAGATGGAG	9109
OY	1501	CACCGGCTCGCGGTGGCGGCGACGTCGAGGAGGGGCTGCGGAGACCCCTTGAGCGCTGCG	1560
Db	9110	CACCGGCTCGCGGTGGCGGCGACGTCGAGGAGGGGCTGCGGAGACCCCTTGAGCGCTGCG	9169
OY	1561	GCGGACGGGACAGACTCGCCCGCGGTGGCGGACGATGCGCGATTCTCAACGGCGGACAG	1620
Db	9170	GCGGACGGGACAGACTCGCCCGCGGTGGCGGACGATGCGCGATTCTCAACGGCGGACAG	9229
OY	1621	CTCGCGCTTCTTTCACCGGACAGGGGGGCGAGACGCTGGGGCATGGGGCGGTGGGCTGTAC	1680
Db	9230	CTCGCGCTTCTTTCACCGGACAGGGGGGCGAGACGCTGGGGCATGGGGCGGTGGGCTGTAC	9289
OY	1681	GATGATATGTCGCGCTTCCGCGAGGCGTTTCGACCTGTGTCGCTGAGGCTGTTCAACAGAG	1740
Db	9290	GATGATATGTCGCGCTTCCGCGAGGCGTTTCGACCTGTGTCGCTGAGGCTGTTCAACAGAG	9349
OY	1741	CTCGACCGGGCGCTTCGCGAGGTGATGTGGGCGGACCGGACAGCGTGCACCGCGCTG	1800
Db	9350	CTCGACCGGGCGCTTCGCGAGGTGATGTGGGCGGACCGGACAGCGTGCACCGCGCTG	9409
OY	1801	CTCGACCGAGACAGCTTCAACCGACGGGGGCGTTTCACCTTGAAATAGGCTGCGCGCG	1860
Db	9410	CTCGACCGAGACAGCTTCAACCGACGGGGGCGTTTCACCTTGAAATAGGCTGCGCGCG	9469
OY	1861	CTGTGGCGGTGCTGTGGGTGTAGAGCCGAGATGTGTCGCGCGCACATAGCATGCTGAGACTG	1920
Db	9470	CTGTGGCGGTGCTGTGGGTGTAGAGCCGAGATGTGTCGCGCGCACATAGCATGCTGAGACTG	9529
OY	1921	GTGGCTGCTCGGTGGCGGGCGGTGTTCTTCGCTTGAGACGCGGCTGTTCTGTGGGCTGCG	1980

[illegible]



Db	8090	CCGAGCGCTGGAGGGGGCCGAATCTCGATATGCTCCCTCGGGGCTTGCGAGGGCCGTGTGTCGC	814.9
Qy	541	GTGATATACGCTCTTATTCGTCTCTCGTGTGGCCGTTTCACTCTGGCTGTCAAGCTTGCGC	600
Db	8150	GTGATATACGCTCTTATTCGTCTCTCGTGTGGCCGTTTCACTCTGGCTGTCAAGCTTGCGC	820.9
Qy	601	TCCGGGGGAATCTCAACGGCCCTGGCTGGTGGGGGATATGCTGATGTGTGTGCCTGAGCAC	660
Db	8210	TCCGGGGGAATCTCAACGGCCCTGGCTGGTGGGGGATATGCTGATGTGTGTGCCTGAGCAC	826.9
Qy	661	CTCGTGTGCTCTCGAAGACCCGGGCGCTGCGCACGAGACGCTGCTGCGAAGCATTTTCG	720
Db	8270	CTCGTGTGCTCTCGAAGACCCGGGCGCTGCGCACGAGACGCTGCTGCGAAGCATTTTCG	832.9
Qy	721	GCGGAGGCCGATGGGTTTCGACGAGGCGAAGGGTGCCTGCTGTGTCTTAAAGCGCTC	780
Db	8330	GCGGAGGCCGATGGGTTTCGACGAGGCGAAGGGTGCCTGCTGTGTCTTAAAGCGCTC	838.9
Qy	781	AGTGGAGCCCGCGGGGACCGCGATTCGGATATTGGCGGGATTCGAGGATCCGGATCAAT	840
Db	8390	AGTGGAGCCCGCGGGGACCGCGATTCGGATATTGGCGGGGATTCGAGGATTCGGATCAAT	844.9
Qy	841	CACGACGGTGGAGACAGCGGCTCTGACCGTGCACAAACGGGAGCTCCCAAGAAATGTCGTG	900
Db	8450	CACGACGGTGGAGACAGCGGCTCTGACCGTGCACAAACGGGAGCTCCCAAGAAATGTCGTG	850.9
Qy	901	AAACGGGCGCTGCGGACGACAGGCTGCGCGCTCTTCGTGGGTATATGTCGAGGCACAC	960
Db	8510	AAACGGGCGCTGCGGACGACAGGCTGCGCGCTCTTCGTGGGTATATGTCGAGGCACAC	856.9
Qy	961	GCGACCGGCGACGACGCTTGTGACCCCATCCGAATCCAACTCTGAAAGCGGTATACGGC	1020
Db	8570	GCGACCGGCGACGACGCTTGTGACCCCATCCGAATCCAACTCTGAAAGCGGTATACGGC	862.9
Qy	1021	CTCGGAGCGAATGTTCGCCACGCGCTGCTGATCCGGTGGGTGAGAACCAACCTTGGCCAT	1080
Db	8630	CTCGGAGCGAATGTTCGCCACGCGCTGCTGATCCGGTGGGTGAGAACCAACCTTGGCCAT	868.9
Qy	1081	CCTGAGTATGCGTGGGAGATCACTGGGCTGCTGAAGGTCTTGTTCCTTCAACACGGG	1140
Db	8690	CCTGAGTATGCGTGGGAGATCACTGGGCTGCTGAAGGTCTTGTTCCTTCAACACGGG	874.9
Qy	1141	CAGATTCCTGGGACCTTCAAGCGCCAGCGCTGAACCCCGGGATCTATGGGGTGAATCTT	1200
Db	8750	CAGATTCCTGGGACCTTCAAGCGCCAGCGCTGAACCCCGGGATCTATGGGGTGAATCTT	880.9
Qy	1201	CGGCTGACCGTCAACGCGCGCCCGGACACCGTGGCGCGGACTGGAATTAACGCCGACGGGCG	1260
Db	8810	CGGCTGACCGTCAACGCGCGCCCGGACACCGTGGCGCGGACTGGAATTAACGCCGACGGGCG	886.9
Qy	1261	GGGGTGAAGCTTCTCGGATGAGCGCGGACCAACCGCACGTTGTGTGGAAGAGCGCGCG	1320
Db	8870	GGGGTGAAGCTTCTCGGATGAGCGCGGACCAACCGCACGTTGTGTGGAAGAGCGCGCG	892.9
Qy	1321	GCGGCGAAGTGCACAACCGCGCGCGCCGGAAGCGACCGGAGAGCTGTGTGTCTGTCCGCA	1380
Db	8930	GCGGCGAAGTGCACAACCGCGCGCGCCGGAAGCGACCGGAGAGCTGTGTGTCTGTCCGCA	898.9
Qy	1381	AGGACCGGCGTGAAGCCCTGGATATGCAACGCGCGCGGGCTGGCGGACATCTGGAAGCCTAC	1440
Db	8990	AGGACCGGCGTGAAGCCCTGGATATGCAACGCGCGCGGGCTGGCGGACATCTGGAAGCCTAC	904.9
Qy	1441	CCTTTCGAGTGTCTGGGCGATGTGGCGTTTCAGTCTGCGGACGAGCGCGCAGCGCATGGAG	1500
Db	9050	CCTTTCGAGTGTCTGGGCGATGTGGCGTTTCAGTCTGCGGAGGAGCGCGCAGCGCATGGAG	910.9
Qy	1501	CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGGCGGCAAGCCCTGGAACGCTGCG	1560
Db	9110	CACCGGCTCGCGGTGGCGGCGACGTCGAGGGAGGGGCTGGCGGCAAGCCCTGGAACGCTGCG	916.9
Qy	1561	GCGGAGGACAGAGCTGCGCCCGCGGTGGCGGTGGCGAGTATCCGCGCAATCTTCAACGCGGCAAG	1620
Db	9170	GCGGAGGACAGAGCTGCGCCCGCGGTGGCGGTGGCGAGTATCCGCGCAATCTTCAACGCGGCAAG	922.9

QY	1621	CTGCGCTTTCTTTCAACCGGACAGGGGGGGCGAGACGCTTGGATATGGGCGCTGGGCTGTAC	1680
Db	9230	CTGGCTTTTCTTTCAACCGGACAGGGGGGGCGAGACGCTTGGGATATGGGCGCTGGGCTGTAC	9289
QY	1681	GATGTATGTGTCGGCGTTCGGCGAGGCGGTTTCGACTGTGTGTAGAGGCTTTCAACGAGAG	1740
Db	9290	GATGTATGTGTCGGCGTTCGGCGAGGCGGTTTCGACTGTGTGTAGAGGCTTTCAACGAGAG	9349
QY	1741	CTGACACCGGCGGCTTCGGCGAGGAGTATGTATGGGGCGAAACGGGCGACGCTGACGCGCGGCTG	1800
Db	9350	CTGACACCGGCGGCTTCGGCGAGGAGTATGTATGGGGCGAAACGGGCGACGCTGACGCGCGGCTG	9409
QY	1801	CTGACACGAGACAGCTTTCAACCCAGCGGCGCTGTTCACTTCGAAATATGAGTGTGCGCGCG	1860
Db	9410	CTGACACGAGACAGCTTTCAACCCAGCGGCGCTGTTCACTTCGAAATATGAGTGTGCGCGCG	9469
QY	1861	CTGTGGCGGCTCTGTGGGGGTATAGAGCGGAGTTGTGTGCGCGCATATAGCATTCGCTATAGCTG	1920
Db	9470	CTGTGGCGGCTCTGTGGGGGTATAGAGCGGAGTTGTGTGCGCGCATATAGCATTCGCTATAGCTG	9529
QY	1921	GTTGAGCTGCGCTGCTGTGGCGGCGGAGTTCTCGCTTGAAGACGCGGATGTTCTGTGTGCTGG	1980
Db	9530	GTTGAGCTGCGCTGCTGTGGCGGCGGAGTTCTCGCTTGAAGACGCGGATGTTCTGTGTGCTGG	9589
QY	1981	CGCGGGGCGCTGATGACGCGGCTGCGGCGCGGCGGCGGCGAGTGTGCTGATCGAGCGCGG	2040
Db	9590	CGCGGGGCGCTGATGACGCGGCTGCGGCGCGGCGGCGGCGAGTGTGCTGATCGAGCGCGG	9649
QY	2041	GAGGCGGAGATGTGGCTCTGCGGTGGGGGGCGGACACGACGCGTGGGTGCTATGCGCGCGGCT	2100
Db	9650	GAGGCGGAGATGTGGCTCTGCGGTGGGGGGCGGACACGACGCGTGGGTGCTATGCGCGCGGCT	9709
QY	2101	AACGCTCCGACCAAGGATGTGATCGCGGGCGCGGCGCAACCCGTGATTCGATTCGCGGCG	2160
Db	9710	AACGCTCCGACCAAGGATGTGATCGCGGGCGCGGCGCGCGCAACCCGTGATTCGATTCGCGGCG	9769
QY	2161	GCGATGCGCGCGCGCGGGGGCGGAAACGAGCGCTTCATGCTTCGATTCGCTTCCACTTCA	2220
Db	9770	GCGATGCGCGCGCGCGGGGGCGGAAACGAGCGCTTCATGCTTCGATTCGCTTCCACTTCA	9829
QY	2221	CCGCTATGAGCCCCGAAATGCTGAGAGCGGTTTCGGGCGGTGTGGCGAGTGCCTGTGACTAACCGG	2280
Db	9830	CCGCTATGAGCCCCGAAATGCTGAGAGCGGTTTCGGGCGGTGTGGCGAGTGCCTGTGACTAACCGG	9889
QY	2281	CGGCGCTGATCGTCTGTGTCAAGCAATCTGAGCGGAGGCTTTCGACAGACGAGGTGAGC	2340
Db	9890	CGGCGCTGATCGTCTGTGTCAAGCAATCTGAGCGGAGGCTTTCGACAGACGAGGTGAGC	9949
QY	2341	TGCGCGGGGCAATTGGGTATGGGCCACCGGCGGAGAGTGTGTGCGGCTCGCGGATGAGGTGAAG	2400
Db	9950	TGCGCGGGGCAATTGGGTATGGGCCACCGGCGGAGAGTGTGTGCGGCTCGCGGATGAGGTGAAG	10009
QY	2401	GCGCTGACGCGCGCGGTGTGGGCACTTTCGTGAGGTGCGTTCGAAATCGACGCTGCTC	2460
Db	10010	GCGCTGACGCGCGCGGTGTGGGCACTTTCGTGAGGTGCGTTCGAAATCGACGCTGCTC	10069
QY	2461	GCGCTGTGCTTCCTTCGTCATGCTCCGAGCGCCCGGCGCGCTGTGCGATTCGTGCGCGCT	2520
Db	10070	GCGCTGTGCTTCCTTCGTCATGCTCCGAGCGCCCGGCGCGCTGTGCGATTCGTGCGCGCT	10129
QY	2521	GGGCGTGAACGACGCGGCGACCGGTCTCGAGAGCGCTGTGGGCGCTGTGGGCGCTCGGTGAGC	2580
Db	10130	GGGCGTGAACGACGCGGCGACCGGTCTCGAGAGCGCTGTGGGCGCTGTGGGCGCTCGGTGAGC	10189
QY	2581	CTGCTCTCTGTGGGCGGCGCTTTCCTTCGAGGGGCGGCGGAGTGCCTGCTGCCAGCTAC	2640
Db	10190	CTGCTCTCTGTGGGCGGCGCTTTCCTTCGAGGGGCGGCGGAGTGCCTGCTGCCAGCTAC	10249
QY	2641	CCCTTGGACGCGAGCGGCTATCTGAGATCGACAGGAAAGCGGACGACGCGGCGGCTGGCGAC	2700
Db	10250	CCCTTGGACGCGAGCGGCTATCTGAGATCGACAGGAAAGCGGACGACGCGGCGGCTGGCGAC	10309

QY 2701 GCGCGTCTCCGAGACGAGGTCACGAGGTCGAGAGGAGGAGGCGCGGTGCGCGCGG 2760  
Db 10310 GCGCGTCTCCGAGACGAGGTCACGAGGTCGAGAGGAGGAGGCGCGGTGCGCGCGG 10369  
QY 2761 GACCGGCGACGCTCGCGCTGACCATCCGCGCGCGCGAGAGGTCGCGCGGAGAGGTC 2820  
Db 10370 GACCGGCGACGCTCGCGCTGACCATCCGCGCGCGCGAGAGGTCGCGCGGAGAGGTC 10429  
QY 2821 GAGCGCGCGAGACCGTCCGTTCCGAGTCGAGTCGATGAGCGAGGCGTCTTATGAC 2880  
Db 10430 GAGCGCGCGAGACCGTCCGTTCCGAGTCGAGTCGATGAGCGAGGCGTCTTATGAC 10489  
QY 2881 CTCTGCTTCCGAGTCACGAGGCGCGCGCGCGCGAGGTCGAGTCGAGTCGAGTCGAGTC 2940  
Db 10490 CTCTGCTTCCGAGTCACGAGGCGCGCGCGCGCGAGGTCGAGTCGAGTCGAGTCGAGTC 10549  
QY 2941 GACCGGCGAGGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTC 3000  
Db 10550 GACCGGCGAGGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTC 10609  
QY 3001 CTGCGCGGAGAGACCAACCTCTCGCTGCTCGAGGCGAGTCGCGCGCGCGCGCGCGTC 3060  
Db 10610 CTGCGCGGAGAGACCAACCTCTCGCTGCTCGAGGCGAGTCGCGCGCGCGCGCGCGTC 10669  
QY 3061 GCGCGGCGAGGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTC 3120  
Db 10670 GCGCGGCGAGGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTC 10729  
QY 3121 GAGAGGTTTGTATCCCATGTCACACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3180  
Db 10730 GAGAGGTTTGTATCCCATGTCACACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10789  
QY 3181 CTCTCGCGATCGAGGCGTC 3240  
Db 10790 CTCTCGCGATCGAGGCGTC 10849  
QY 3241 GACAGATAGTCCCGCTTCAAGCGCGGAGCGCGGTCGATTCATGCGCGAGCGCGCGG 3300  
Db 10850 GACAGATAGTCCCGCTTCAAGCGCGGAGCGCGGTCGATTCATGCGCGAGCGCGCGG 10909  
QY 3301 GTCCGTCTCG 3360  
Db 10910 GTCCGTCTCG 10969  
QY 3361 GGCAGCGCGCGAGAAACCGCGCTTCACTGAGTCGTCGCGCGCGCGCGCGCGCGCGCGCGCG 3420  
Db 10970 GGCAGCGCGCGAGAAACCGCGCTTCACTGAGTCGTCGCGCGCGCGCGCGCGCGCGCGCG 11029  
QY 3421 CGCTCGAGCGCGGTCGTCG 3480  
Db 11030 CGCTCGAGCGCGGTCGTCG 11089  
QY 3481 GTGCTCAACTGCTCTCGCGCGAGTCGATTCGACAAAGATTTCATCTCTCGCGAGTCGAC 3540  
Db 11090 GTGCTCAACTGCTCTCGCGCGAGTCGATTCGACAAAGATTTCATCTCTCGCGAGTCGAC 11149  
QY 3541 GGCAGGTTTGTGAGCTCGAGCGAGCGAGCTGTAACGCGAGTAACCACTCGCGCGTCGCG 3600  
Db 11150 GGCAGGTTTGTGAGCTCGAGCGAGCGAGCTGTAACGCGAGTAACCACTCGCGCGTCGCG 11209  
QY 3601 CGGTCCTGCGCAATCT 3660  
Db 11210 CGGTCCTGCGCAATCT 11269  
QY 3661 CCGCGCGCGGTCGTCGCGCTCTTGGAGAGCTCTCTCGCGCTGATTCGCGCGAGCGGTTC 3720  
Db 11270 CCGCGCGCGGTCGTCGCGCTCTTGGAGAGCTCTCTCGCGCTGATTCGCGCGAGCGGTTC 11279  
QY 3721 ACCCTCCCGCATCGAGCGCTCCGATTCGCGCGGTCGCGCGAGCGGTTCGCGAGCATG 3780  
Db 11330 ACCCTCCCGCATCGAGCGCTCCGATTCGCGCGGTCGCGCGAGCGGTTCGCGAGCATG 11389  
QY 3781 GCGCAGGCGCAGCATCTTGGAGAGCTCGTACTCACTGTCGTCGTCGTCGTCGTCGTCGTCG 3840

Db 11390 GCGCAGGCGCAGCATCTTGGAGAGCTCGTACTCACTGTCGTCGTCGTCGTCGTCGTCGTCG 11449  
QY 3841 CGTATTCACACCGAGCG 3900  
Db 11450 CGTATTCACACCGAGCG 11509  
QY 3901 CTCCGCTGAGCTGCG 3960  
Db 11510 CTCCGCTGAGCTGCG 11569  
QY 3961 GTCTCGAGGTCGTCG 4020  
Db 11570 GTCTCGAGGTCGTCG 11629  
QY 4021 CTGCGCATGAGCTGCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4080  
Db 11630 CTGCGCATGAGCTGCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11689  
QY 4081 CTGAGCTGTCGAGCGAGCTTCTGTCGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4140  
Db 11690 CTGAGCTGTCGAGCGAGCTTCTGTCGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11749  
QY 4141 CTGTTGATGCTCTGCG 4200  
Db 11750 CTGTTGATGCTCTGCG 11809  
QY 4201 GCAGCGCGTGCAGAAACGACTTCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4260  
Db 11810 GCAGCGCGTGCAGAAACGACTTCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11869  
QY 4261 CTATGA 4266  
Db 11870 CTATGA 11875

RESULT 5  
ARI99567 68750 bp DNA linear PAT 20-APR-2002  
LOCUS ARI99567  
DEFINITION Sequence 1 from patent US 6355459.  
ACCESSION ARI99567  
VERSION ARI99567.1 GI:20249641  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 68750)  
AUTHORS Schnupp,T., Iijon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.  
TITLE Genes for the biosynthesis of epoethilones  
JOURNAL Patent: US 6355459-A, 1 12-MAR-2002;  
FEATURES  
source Location/Qualifiers  
1..68750  
/organism="unknown"  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN

Query Match 100.0%; Score 4266; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pired. No. 0;  
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGGATGCTGCTCCATCGAGCGCGAGCGGAGATCCGATTGCGATGTCGAGCGAGT 60  
Db 7610 GTGCGGATGCTGCTCCATCGAGCGCGAGCGGAGATCCGATTGCGATGTCGAGCGAGT 7669  
QY 61 TGCGGTCTGCGCGGTCGTCGATGATCTGAGCGGCTTCTGAGCGCTCTCGAGCGCTCG 120  
Db 7670 TGCGGTCTGCGCGGTCGTCGATGATCTGAGCGGCTTCTGAGCGCTCTCGAGCGCTCG 7729  
QY 121 CGCAGCACGTCGCGCGAGTCGCGCGCGAGAGCTGCGAGATGACAGAGCTGTTGATCC 180  
Db 7730 CGCAGCACGTCGCGCGAGTCGCGCGCGAGAGCTGCGAGATGACAGAGCTGTTGATCC 7789

OY	1.81	AACCCCGATGGCCCGGGGAAACAGCCGGCTPACGGCGCATCTTTTCTGAGGAGGCTAGCC	240
Db	7790	GACCCCGATGGCCCGGGGAAAGCCCGCTTACGGCGCATCTTTCTGAGGAGGCTAGCC	7849
OY	241	TGCTTCGACGCGCTCTCTTCGGGATCTCGGCTCGGAGCGCTCGGATGGAACCTTGCA	300
Db	7850	TGCTTCGACGCGCTCTCTTCGGGATCTCGGCTCGGAGCGCTCGGATGGAACCTTGCA	7909
OY	301	CATCGACTCTTGTGTGAGGTGTGCTGGAGCGCTGAGAACCGCGCGATGCTTCATCG	360
Db	7910	CATCGACTCTTGTGTGAGGTGTGCTGGAGCGCTGAGAACCGCGCGATGCTTCATCG	7969
OY	361	GCGCTCCGTGGGTACGGAAACGGGAGTGTTCATTCGGGATCGGCGCTCCGATATGAGGCC	420
Db	7970	GCGCTCCGTGGGTACGGAAACGGGAGTGTTCATTCGGGATCGGCGCTCCGATATGAGGCC	8029
OY	421	GCGCTCCGGCAAGCGACGGCGCTCGCAGATTCGACGCTCATGTGGCGGGCTGGGACGATG	480
Db	8030	GCGCTCCGGCAAGCGACGGCGCTCGCAGATTCGACGCTCATGTGGCGGGCTGGGACGATG	8089
OY	481	CCGACCGTCGAGCGGGGCGGAATCTCGTATSCCTTCGGGCTTGCAAGGCGGTGTGTGCGG	540
Db	8090	CCGACCGTCGAGCGGGGCGGAATCTCGTATSCCTTCGGGCTTGCAAGGCGGTGTGTGCGG	8149
OY	541	GTGGAATAGGCGCTATTCGTCTCTCGTGTGACCGCTTCATCTGGCTGTCAAGCTTGTGGC	600
Db	8150	GTGGAATAGGCGCTATTCGTCTCTCGTGTGACCGCTTCATCTGGCTGTCAAGCTTGTGGC	8209
OY	601	TCCGGGGAATCTCCACGGGCGCTGGCTGTGGGGGATATCGCTGATGTTGTTCGCCAGCAC	660
Db	8210	TCCGGGGAATCTCCACGGGCGCTGGCTGTGGGGGATATCGCTGATGTTGTTCGCCAGCAC	8269
OY	661	CTCGTGTGGCTCTCGAAGACCCGAGCGCTGCGCAGAGACGATCGCTGCAAGGCAATTTTCG	720
Db	8270	CTCGTGTGGCTCTCGAAGACCCGAGCGCTGCGCAGAGACGATCGCTGCAAGGCAATTTTCG	8329
OY	721	GCGGAGGCGCCATATGGTATTTCGACAGAGGCGAAGGATCGCGCTGTGTGTCTTCAACCGGCTC	780
Db	8330	GCGGAGGCGCCATATGGTATTTCGACAGAGGCGAAGGATCGCGCTGTGTGTCTTCAACCGGCTC	8389
OY	781	AGTGAGAGCCCGCGCGGACGCGGATGGATATTTGGCGGTGATTTCGAGATTCGCGCATCAT	840
Db	8390	AGTGAGAGCCCGCGCGGACGCGGATGGATATTTGGCGGTGATTTCGAGATTCGCGCATCAT	8449
OY	841	CACGACGSGTGCAGACGAGCGGCTGTGACCGTGCGCCAACGGGAGCTCCCAAGAAATCGTGTG	900
Db	8450	CACGACGSGTGCAGACGAGCGGCTGTGACCGTGCGCCAACGGGAGCTCCCAAGAAATCGTGTG	8509
OY	901	AAACGGGCGCTTGTGGGACGCGACGCTGTGCGCGCTTCTTGTGTGGGTATATGTGAAGCACAC	960
Db	8510	AAACGGGCGCTTGTGGGACGCGACGCTGTGCGCGCTTCTTGTGTGGGTATATGTGAAGCACAC	8569
OY	961	GGCACGGGCAAGACGCTTGATGATCCCAATTCGAATTCGAAGCTCGAATGCGGTATAGCGG	1020
Db	8570	GGCACGGGCAAGACGCTTGATGATCCCAATTCGAATTCGAAGCTCGAATGCGGTATAGCGG	8629
OY	1021	CTCGGCGGAGATGTGCCACGCGCGCTGTGATCGGGTCGATGAAGCAACCTTGGCCAT	1080
Db	8630	CTCGGCGGAGATGTGCCACGCGCGCTGTGATCGGGTCGATGAAGCAACCTTGGCCAT	8689
OY	1081	CTGTAGTATGTGTCGGGGATCATCTGTGGCTGTGTAAGGTGTCTTCCCTTCAGACGCGG	1140
Db	8690	CTGTAGTATGTGTCGGGGATCATCTGTGGCTGTGTAAGGTGTCTTCCCTTCAGACGCGG	8749
OY	1141	CAGATTTCTCGGCACTCCACGCGCGGAGGCGCTGAACCCCGGATTCATGAGGGGTGATTT	1200
Db	8750	CAGATTTCTCGGCACTCCACGCGCGGAGGCGCTGAACCCCGGATTCATGAGGGGTGATTT	8809
OY	1201	CGGCTGACCTGTACCGCGCGCCCGGACACCGGTGCGCGACTGGAATACCCCGCGACGCGCG	1260
Db	8810	CGGCTGACCTGTACCGCGCGCCCGGACACCGGTGCGCGACTGGAATACCCCGCGACGCGCG	8869
OY	1261	GCGGTGACCTGTTCCGCGATGAGCGGAGCAACGCGCACGTGTGTCTGGAAGAGCGCGC	1320

Db	8870	GGGGTAGGCTGTTCCGGCAATGAGGGGAGCCAAAGCCGACGTGGTGGCTGGAAGAGCGCCG	8929
Oy	1321	GGGGCGACGTGACACCGCCGGCGCGGAGCGACCGGACAGGCTGCTGGTCTGTGCGCA	1380
Db	8930	GGGGCGACGTGACACCGCCGGCGCGCGAGCGACCGGCAGAGCTGGTGGTCTGTGCGCA	8989
Oy	1381	AGGACCCGCTCAGCCCTGAGATGACAGAGGGGGCGCGTGGGGAGACCATCTGGAGACCTAC	1440
Db	8990	AGGACCCGCTCAGCCCTGAGATGACAGAGGGGGCGCGCGTGGGGAGACCATCTGGAGACCTAC	9049
Oy	1441	CCCTTCGAGAGTGTGGGGCGATGTGGCGCTTCACTGTGGCGACGACGGCGACGGATGGAG	1500
Db	9050	CCCTTCGAGAGTGTGGGGCGATGTGGCGCTTCACTGTGGCGACGACGGCGACGGATGGAG	9109
Oy	1501	CACCGGCTCGCGGTGGCGGAGCGTCGAGGGAGGGGGCTGGGGCAGCCCTGGACGCTGCG	1560
Db	9110	CACCGGCTCGCGGTGGCGGAGCGTCGAGGGAGGGGGCTGGGGCAGCCCTGGACGCTGCG	9169
Oy	1561	GGCGAGGGAGACGAGCTGCCCGGGTGGCGGTGCGGATTCGCGCATTCCTACCGCGGAG	1620
Db	9170	GGCGAGGGAGACGAGCTGCCCGGGTGGCGGTGCGGATTCGCGCATTCCTACCGCGGAG	9229
Oy	1621	CTGGCTTTCTCTTCAACCGGACAGGGGGGCGCAGACGCTGGGCAATGGGGCGTGGGCTGTAC	1680
Db	9230	CTGGCTTTCTCTTCAACCGGACAGGGGGGCGCAGACGCTGGGCAATGGGGCGTGGGCTGTAC	9289
Oy	1681	GATGTATGTGTCGGCGTTCGCGAGGCGCTTGACCTGTGCTGAGGCTGTTCAACAGAG	1740
Db	9290	GATGTATGTGTCGGCGTTCGCGAGGCGCTTGACCTGTGCTGAGGCTGTTCAACAGAG	9349
Oy	1741	CTGCAACGGCGCGCTCCGCGAGGTGATGTGGGCGGACACCGGCAAGCTGACAGCGCGCTG	1800
Db	9350	CTGCAACGGCGCGCTCCGCGAGGTGATGTGGGCGGACACCGGCAAGCTGACAGCGCGCTG	9409
Oy	1801	CTGCAACGAGCGCTTCAACCGGCGCGCGCTGTTCACCTTCGAAATATGGCGTCCGCGCG	1860
Db	9410	CTGCAACGAGCGCTTCAACCGGCGCGCGCTGTTCACCTTCGAAATATGGCGTCCGCGCG	9469
Oy	1861	CTGTGGCGGTGCTGGGGGTGTAGAGCCCGAGTTGGTCGCGCGGCCAATAGCATCGGTGACTG	1920
Db	9470	CTGTGGCGGTGCTGGGGGTGTAGAGCCCGAGTTGGTCGCGCGGCCAATAGCATCGGTGACTG	9529
Oy	1921	GTGGCTGCTGCGTGGCGGGCGGTGTTCCGCTTGAGGACCGCGGTGTTCTGTGGTGGCTGCG	1980
Db	9530	GTGGCTGCTGCGTGGCGGGCGGTGTTCCGCTTGAGGACCGCGGTGTTCTGTGGTGGCTGCG	9589
Oy	1981	CGCGGGGCGCGATGACAGGGCGCTGCGCGCGCGGGCGGAGTGGTGTGATCTGAGAGCGCG	2040
Db	9590	CGCGGGGCGCGATGACAGGGCGCTGCGCGCGCGGGCGGAGTGGTGTGATCTGAGAGCGCG	9649
Oy	2041	GAGGCGGATGTGGCTGCTGGGTGGCGCGCACGACGTCGGGTGTGATCGCCCGCGGCTC	2100
Db	9650	GAGGCGGATGTGGCTGCTGGGTGGCGCGCACGACGTCGGGTGTGATCGCCCGCGGCTC	9709
Oy	2101	AAGCCTCCGACCAAGGTGTGATATGCGGGGGCGCGGGGCAACCCGTGTGATGTGATCTGCGCG	2160
Db	9710	AAGCCTCCGACCAAGGTGTGATATGCGGGGGCGCGGGGCAACCCGTGTGATGTGATCTGCGCG	9769
Oy	2161	GGGATGGCGGGCGGGGGGCGGAAACAAAGCGCTCACGCTTCGACATGTGCTTCATCTCA	2220
Db	9770	GGGATGGCGGGCGGGGGGCGGAAACAAAGCGCTCACGCTTCGACATGTGCTTCATCTCA	9829
Oy	2221	CCGCTCATGTGGCCCGATGTGAGCGCTTGGGGCGTGGGCGGATCGGATCGGTGAGCTACCG	2280
Db	9830	CCGCTCATGTGGCCCGGATGTGAGCGCTTGGGGCGTGGGCGGATCGGATCGGTGAGCTACCG	9889
Oy	2281	CGGCGCTGTGATGTTCTGTGTGACGAACTGTAGCGGGAGAGCTTGTGCACAGACGAGGTGAC	2340
Db	9890	CGGCGCTGTGATGTTCTGTGTGACGAACTGTAGCGGGAGAGCTTGTGCACAGACGAGGTGAC	9949
Oy	2341	TGCGCGGGCTATTTGGGTGGCGCAAGCGCGAGAGGTGGTGGGCTTCGGGATGTGAGTGAAG	2400





FEATURES Location/Qualifiers  
source 1..68750  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN

Query Match 100.0%; Score 4266; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGGCGATGTCCTCCATCGAGCGCGAGCCGAAGATCCGATTGCGATCGTCGAGCGAGT 60  
Db GTGGCGATGTCCTCCATCGAGCGCGAGCCGAAGATCCGATTGCGATCGTCGAGCGAGT 7669

61 TCCCGTCTGCCGCTGCGTGTGATCGATCTGAGCGGGTCTGGAACGCTCTTCAAGCGCTCG 120  
Db TCCCGTCTGCCGCTGCGTGTGATCGATCTGAGCGGGTCTGGAACGCTCTTCAAGCGCTCG 7729

121 CGCGACCGCTGGGGAGTCCCGCGGAAAGCTGGGATGCAAGCGCTGGTTGATCC 180  
Db CGCGACCGCTGGGGAGTCCCGCGGAAAGCTGGGATGCAAGCGCTGGTTGATCC 7789

181 GACCCGATGCCCGGGGAAAGAGCCGTTACGCGCGCATCTTCTCTGACGACGTAAGC 240  
Db GACCCGATGCCCGGGGAAAGAGCCGTTACGCGCGCATCTTCTCTGACGACGTAAGC 7849

241 TGCCTGACGCTCTCTTCTGCGCATCTGCGCTCGGAAAGCGCTGCGAATGGAACCTTGA 300  
Db TGCCTGACGCTCTCTTCTGCGCATCTGCGCTCGGAAAGCGCTGCGAATGGAACCTTGA 7909

301 CATTCAGCTTGTGAGGATGCTGGGAGGCGCTGGAGAAAGCGCGCATGCTCCATCG 360  
Db CATTCAGCTTGTGAGGATGCTGGGAGGCGCTGGAGAAAGCGCGCATGCTCCATCG 7969

361 GCGCTGTCGTGATCGGAAACGGGAGTGTTCATCGGATCGCGCGCTCGGATATGAGGC 420  
Db GCGCTGTCGTGATCGGAAACGGGAGTGTTCATCGGATCGCGCGCTCGGATATGAGGC 8029

421 GCGCTGCGGAAAGCGAGCGCGCTCGGAGATGACGCTCATGAGCGGCTGGGAGCATG 480  
Db GCGCTGCGGAAAGCGAGCGCGCTCGGAGATGACGCTCATGAGCGGCTGGGAGCATG 8089

481 CCCAGGTCGAGGCGGCGGAGATCTGTATGCTTCGCGGCTGGAGGCGGTGTGCGG 540  
Db CCCAGGTCGAGGCGGCGGAGATCTGTATGCTTCGCGGCTGGAGGCGGTGTGCGG 8149

541 GTGATATCGGCTTATCTGCTCGCTGATGAGCGTTCATCTGCGCTGTGAGAGCTTGGC 600  
Db GTGATATCGGCTTATCTGCTCGCTGATGAGCGTTCATCTGCGCTGTGAGAGCTTGGC 8209

601 TCCGGGAAATGCTCCACGCGCTGCTGAGGATTCGCTGATGTTGTGCGGACAC 660  
Db TCCGGGAAATGCTCCACGCGCTGCTGAGGATTCGCTGATGTTGTGCGGACAC 8269

661 CTCGTGTGCTCTCGAAGACCGGGGCGTGGCCAGGAGCGTTCGCGCAAGGATTTTG 720  
Db CTCGTGTGCTCTCGAAGACCGGGGCGTGGCCAGGAGCGTTCGCGCAAGGATTTTG 8329

721 GCGGAGCGGATGAGTTCGAGCGAGGCGAAGGAGTGGCGCTGTGTCTTCAAGCGGCTC 780  
Db GCGGAGCGGATGAGTTCGAGCGAGGCGAAGGAGTGGCGCTGTGTCTTCAAGCGGCTC 8389

781 AGTGAAGCCCGCGGAGCGGATCGGATATTTGCGGTATTCGAGGATCCGCGATCAAT 840  
Db AGTGAAGCCCGCGGAGCGGATCGGATATTTGCGGTATTCGAGGATCCGCGATCAAT 8449

841 CACGACGCTGCGAGCGAGCTGACACGTCGCGAAGCGGAGCTCCCAAGAAATCGGCTG 900  
Db CACGACGCTGCGAGCGAGCTGACACGTCGCGAAGCGGAGCTCCCAAGAAATCGGCTG 8509

901 AAACGGGCTTGGCGAGCGCAGGCTGCGCGCGTCTTCTGAGGATATGTCAGGACAC 960  
Db AAACGGGCTTGGCGAGCGCAGGCTGCGCGCGTCTTCTGAGGATATGTCAGGACAC 8569

961 GGCACGGGACAGACGCTTGTGATCCCATTCGAATCCAGCTTGAATGCGGTATACGGC 1020  
Db GGCACGGGACAGACGCTTGTGATCCCATTCGAATCCAGCTTGAATGCGGTATACGGC 8629

1021 CTCGGGACGATGTCGACCGCGCTGCTGATTCGGGTGCTGAGAACCAACTTGGCCAT 1080  
Db CTCGGGACGATGTCGACCGCGCTGCTGATTCGGGTGCTGAGAACCAACTTGGCCAT 8689

1081 CTTGATATGCTGCGGGATCACTGGGCTGTGAAGTGTCTTCTTCCCTTACGACGGG 1140  
Db CTTGATATGCTGCGGGATCACTGGGCTGTGAAGTGTCTTCTTCCCTTACGACGGG 8749

1141 CAGATTCCTGGGACCTTCCACGCGAGCGCTGAAACCCCGGATTCATGAGGATGCTT 1200  
Db CAGATTCCTGGGACCTTCCACGCGAGCGCTGAAACCCCGGATTCATGAGGATGCTT 8809

1201 CGGCTGACCGTACCGCGCGCCCGGACACCTGTGCGCGACTGGAATACGCGGAGCG 1260  
Db CGGCTGACCGTACCGCGCGCCCGGACACCTGTGCGCGACTGGAATACGCGGAGCG 8869

1261 GGGGTGACCTGCTTGGCATGACGCGGACCAACGCGACGTGTGCTGAAAGGCGCG 1320  
Db GGGGTGACCTGCTTGGCATGACGCGGACCAACGCGACGTGTGCTGAAAGGCGCG 8929

1321 GCGGACGATGACACCGCGCGCGCGCGGACGACCGGAGAGCTGTGCTGCGGCA 1380  
Db GCGGACGATGACACCGCGCGCGCGCGGACGACCGGAGAGCTGTGCTGCGGCA 8989

1381 AGGACCGGTCAGACCTGATGACAGGCGCGCGCTGCGGACCATCTGAGACCTTAC 1440  
Db AGGACCGGTCAGACCTGATGACAGGCGCGCGCTGCGGACCATCTGAGACCTTAC 9049

1441 CTTTGCAGTGTCTGGGCGATGTGCGTTCAGTCTGCGACGACGCGGACGCGATGAG 1500  
Db CTTTGCAGTGTCTGGGCGATGTGCGTTCAGTCTGCGACGACGCGGACGCGATGAG 9109

1501 CACCGGCTCGGAGTGGCGGCGGACCGTGCAGGAGGCGGCGCGGACCGCTGAGCGCTCG 1560  
Db CACCGGCTCGGAGTGGCGGCGGACCGTGCAGGAGGCGGCGGACCGCTGAGCGCTCG 9169

1561 GCGCAGGACAGACGTCGCCGCGTGGTGGCGAGTATGCGGATTCCTCAACGCGGACG 1620  
Db GCGCAGGACAGACGTCGCCGCGTGGTGGCGAGTATGCGGATTCCTCAACGCGGACG 9229

1621 CTCGCTTCTTCTTCAACGCGGACGCGGCGGACGCTGGGATGAGCCGCTGCTTAC 1680  
Db CTCGCTTCTTCTTCAACGCGGACGCGGCGGACGCTGGGATGAGCCGCTGCTTAC 9289

1681 GATGTATGTCGCGCTTCCGAGGCGGTCGACCTGTGCGGTGAGGCTGTTCAACGAGAG 1740  
Db GATGTATGTCGCGCTTCCGAGGCGGTCGACCTGTGCGGTGAGGCTGTTCAACGAGAG 9349

1741 CTCGACCGGCGCTCGGAGGATGTGAGCGGACGCGGACGCGGACGCGGCTGCGCTG 1800  
Db CTCGACCGGCGCTCGGAGGATGTGAGCGGACGCGGACGCGGACGCGGCTGCGCTG 9409

1801 CTCGACGAGACGCTTCAACGCGGCGGCTGTTCACTTCAATATGCGCTCCCGCG 1860  
Db CTCGACGAGACGCTTCAACGCGGCGGCTGTTCACTTCAATATGCGCTCCCGCG 9469

1861 CTCGTGCGGCTGTGGGATGTGAGCGGAGGATGTGCGGCGGACATGAGATGAGGCTG 1920  
Db CTCGTGCGGCTGTGGGATGTGAGCGGAGGATGTGCGGCGGACATGAGATGAGGCTG 9529

1921 GTGCTGCTGCGGCTGCGGCGCTTCTGCTTGAAGACGCGGCTTCTGCTGCTGCG 1980  
Db GTGCTGCTGCGGCTGCGGCGCTTCTGCTTGAAGACGCGGCTTCTGCTGCTGCG 9589

1981 CGCGGCGCTGATGACGCGCTGCGCGCGGCGGCGGATGTGATGAGGCGCG 2040  
Db CGCGGCGCTGATGACGCGCTGCGCGCGGCGGCGGATGTGATGAGGCGCG 9649



Qy	2041	GAGGCGCAATGAGGTGCTGCGAGTGGCCGCCGACACAGCTCGGTGTGATATCGCCGCGGTC	2100
Dp	9650	GAGGCGCAATGAGGTGCTGCGAGTGGCCGCCGACACAGCTCGGTGTGATATCGCCGCGGTC	9709
Qy	2101	AACGCTCCGGACCAAGTGTGATATGCGGAGGCCCGGGGCAACCCGGTCATGTGATCGAGCG	2160
Dp	9710	AACGCTCCGGACCAAGTGTGATATGCGGAGGCCCGGGGCAACCCGGTCATGTGATCGAGCG	9769
Qy	2161	GCGATGCGCGGCGCGGGGCGCGAACAAGCGCTCCAGCTCTTCGACATGCGTTCCACTCA	2220
Dp	9770	GCGATGCGCGGCGCGGGGCGCGAACAAGCGCTCCAGCTCTTCGACATGCGTTCCACTCA	9829
Qy	2221	CCGCTCATATGCCCCCGATATCTGTGAAGCGTTGAGGCGTGTGGCCGAGTGGAGCTACCG	2280
Dp	9830	CCGCTCATATGCCCCCGATATCTGTGAAGCGTTGAGGCGTGTGGCCGAGTGGAGCTACCG	9889
Qy	2281	CGGCGGTCGATCGATCTGTGAGCAATCTGAGCGGAAAGGCTTGACAGACGAGGTGAGC	2340
Dp	9890	CGGCGGTCGATCGATCTGTGAGCAATCTGAGCGGAAAGGCTTGACAGACGAGGTGAGC	9949
Qy	2341	TCGCGGAGCTAATGGATGCGCCACAGCGCGAGAGGTGTGTGCGTTGCGGATGAGATGAAG	2400
Dp	9950	TCGCGGAGCTAATGGATGCGCCACAGCGCGAGAGGTGTGTGCGTTGCGGATGAGATGAAG	10000
Qy	2401	GCGCTGACACGCGCGCGGTGCGGGGACCTTTGTGAGTCCGTCCGAATGACGCTGCTC	2460
Dp	10010	GCGCTGACACGCGCGCGGTGCGGGGACCTTTGTGAGTCCGTCCGAATGACGCTGCTC	10065
Qy	2461	GGCGCTGAGGCTGCTGTCATACGCGGAGAGCGCCGCGCGGCGCTGCTGTCGATGTCGCGGCT	2520
Dp	10070	GGCGCTGAGGCTGCTGTCATACGCGGAGAGCGCCGCGCGGCGCTGCTGTCGATGTCGCGGCT	10122
Qy	2521	GAGCGGTGACGAGCCGCGGACCGGTGCTGAGAGCGCTCGGCGGAGCTTGAGGCGCTGAGTGC	2580
Dp	10130	GAGCGGTGACGAGCCGCGGACCGGTGCTGAGAGCGCTCGGCGGAGCTTGAGGCGCTGAGTGC	10188
Qy	2581	CTGGTCTCTGTGGGCGCGCTCTTTCCCTCAGAGGGGCGGCGGAGTCCGCTGCCACGTAC	2640
Dp	10190	CTGGTCTCTGTGGGCGCGCTCTTTCCCTCAGAGGGGCGGCGGAGTCCGCTGCCACGTAC	10244
Qy	2641	CCTTGGGACGCGGAGCGCTACTGATATGACACGAAACCGACGAGCGGCGCGTGGGAGC	2700
Dp	10250	CCTTGGGACGCGGAGCGCTACTGATATGACACGAAACCGACGAGCGGCGCGTGGGAGC	10305
Qy	2701	CGCGGTGCTCCGAGAGCGGATCAAGACGAGTTCGAGAGAGGGGAGCGGATGCGCGAGCGC	2760
Dp	10310	CGCGGTGCTCCGAGAGCGGATCAAGACGAGTTCGAGAGAGGGGAGCGGATGCGCGAGCGC	10366
Qy	2761	GACCGGCGACGCGTTCGAGCTGACCATTCGCGCGCCGACGAGCGGACGCGCGGAGAAAGTC	2820
Dp	10370	GACCGGCGACGCGTTCGAGCTGACCATTCGCGCGCCGACGAGCGGACGCGCGGAGAAAGTC	10422
Qy	2821	GAGGCGCGCGGCGACCGTCCGTTCCGGTTCGAGATGATAGGCCAGGCGGTGTGATAC	2880
Dp	10430	GAGGCGCGCGGCGACCGTCCGTTCCGGTTCGAGATGATAGGCCAGGCGGTGTGATAC	10485
Qy	2881	CTCGTGCTTGGGATCAAGAGCGGCGCGCCCTGTGTCTTGGGCGAGGTGAGATTCGCGTCT	2940
Dp	10490	CTCGTGCTTGGGATCAAGAGCGGCGCGCCCTGTGTCTTGGGCGAGGTGAGATTCGCGTCT	10544
Qy	2941	GACGCGGCGGAGCTCAGCTTCATATGATGTCCAGTTCGCGCTGGGCGATGTGCCGACGAC	3000
Dp	10550	GACGCGGCGGAGCTCAGCTTCATATGATGTCCAGTTCGCGCTGGGCGATGTGCCGACGAC	10605
Qy	3001	CTGCGCGGAAAGCCCAACCTTCGCGCTGTGTCCGAGAGCGAGTGGCGCGGCGCATGTCTC	3060
Dp	10610	CTGCGCGGAAAGCCCAACCTTCGCGCTGTGTCCGAGAGCGAGTGGCGCGGCGCATGTCTC	10665
Qy	3061	GCGCTGGGCGAGGCGGTGAAAGGCTCTGTGGTGGGCAACCGGTCATTCGCGCTTTGGGCG	3120
Dp	10670	GCGCTGGGCGAGGCGGTGAAAGGCTCTGTGGTGGGCAACCGGTCATTCGCGCTTTGGGCG	10722
Qy	3121	GGAGCGTTTGTCAACCACTCACCAAGTCGGCTGCGCTGAGTCTCGGCTCAAGGCG	3180

[illegible]

Db 11810 GCAGGCGTCGACAAAGACGACTTGTCTTCATCGGGGCGACATCAAGACTGGAAATCATTTGCC 11869  
Qy 4261 CTATGA 4266  
Db 11870 CTATGA 11875

RESULT 7  
AR208671  
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 1 from patent US 6383787.  
ACCESSION AR208671  
VERSION AR208671.1 GI:21509886  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
AUTHORS 1 (bases 1 to 68750)  
Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and  
Goriach,J.  
TITLE Genes for the biosynthesis of epochilones  
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;  
FEATURES  
source 1..68750  
/organism="Unknown"  
BASE COUNT 9596 a 22456 c 25539 g 11159 t  
ORIGIN

Query Match 100.0%; Score 4266; DB 6; Length 68750;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGGGATCGTCCCATCGACGGCCGACGCCGAAGATCCGATTGCCATGCTGGACGAGT 60  
Db 7610 GTGGGGGATCGTCCCATCGACGGCCGACGCCGAAGATCCGATTGCCATGCTGGACGAGT 7669  
Qy 61 TGCCTGTGCGCCGCTGCGATGATCGATCGATGAGCGGGTTCTGACGCTCTCGAGGGCTCG 120  
Db 7670 TGCCTGTGCGCCGCTGCGATGATCGATCGATGAGCGGGTTCTGACGCTCTCGAGGGCTCG 7729  
Qy 121 CGCGACACCGTGGGCGAGTCCCGCGCAACGCTGGGATCGACAGCGTGTGATGCC 180  
Db 7730 CGCGACACCGTGGGCGAGTCCCGCGCAACGCTGGGATCGACAGCGTGTGATGCC 7789  
Qy 181 GACCCCGATGCCCCGGGGAAGACGCCCTTACGCGCGCATCTTTCTGAGCGATGAGCC 240  
Db 7790 GACCCCGATGCCCCGGGGAAGACGCCCTTACGCGCGCATCTTTCTGAGCGATGAGCC 7849  
Qy 241 TGCCTGACGCGCTCTTCTTTCGACATCTGCGCTCGGAAAGCGCTGGATGAGCCCTGCA 300  
Db 7850 TGCCTGACGCGCTCTTCTTTCGACATCTGCGCTCGGAAAGCGCTGGATGAGCCCTGCA 7909  
Qy 301 CATCGACTTGTCTGAGAGGTGTGCTGGAGGCGCTGGAGAACGCGCGATGCTCCATG 360  
Db 7910 CATCGACTTGTCTGAGAGGTGTGCTGGAGGCGCTGGAGAACGCGCGATGCTCCATG 7969  
Qy 361 GGGCTGCTGCGATCGGAAACGGGAGGTTCATCGGATGAGCGCCGCTCGGAATATAGGCC 420  
Db 7970 GGGCTGCTGCGATCGGAAACGGGAGGTTCATCGGATGAGCGCCGCTCGGAATATAGGCC 8029  
Qy 421 GCGCTGCGCAAGCAGCGCGCTGCGCAGAGATCGAGCTCATGCGCGGCTGCGGAGAGT 480  
Db 8030 GCGCTGCGCAAGCAGCGCGCTGCGCAGAGATCGAGCTCATGCGCGGCTGCGGAGAGT 8089  
Qy 481 CCAGCGTTCGAGAGCGGGCGGAATCTGTATGCTTCGAGGCTCGAGAGGCGGTGTGCGG 540  
Db 8090 CCAGCGTTCGAGAGCGGGCGGAATCTGTATGCTTCGAGGCTCGAGAGGCGGTGTGCGG 8149  
Qy 541 GTGAGTACGGCTTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Db 8150 GTGAGTACGGCTTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8209  
Qy 601 TCCGGGGAATGCTCCACGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

Db 8210 TCCGGGGAATGCTCCACGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8269  
Qy 661 CTGCTGTGCTCTCTGAGAACCCCGGCGCTGCGCAGAGACGCTGCTGCAAGGCATTTTCG 720  
Db 8270 CTGCTGTGCTCTCTGAGAACCCCGGCGCTGCGCAGAGACGCTGCTGCAAGGCATTTTCG 8329  
Qy 721 GCGAGGCGCATGAGTTGAGACGAGGCGAAGGTCGCGCTGCTGCTGCTGCTGCTGCTGCTG 780  
Db 8330 GCGAGGCGCATGAGTTGAGACGAGGCGAAGGTCGCGCTGCTGCTGCTGCTGCTGCTGCTG 8389  
Qy 781 AGTGAAGCCCGCGCGAGACGCGCATGAGTATTTGGCGGATTCAGAGATCCGCGATCAT 840  
Db 8390 AGTGAAGCCCGCGCGAGACGCGCATGAGTATTTGGCGGATTCAGAGATCCGCGATCAT 8449  
Qy 841 CACGACGCTGCGAGCAGCGGTCTGACCGTGTCCGAAACGAGAGCTCCCAAGAAATGCTGCTG 900  
Db 8450 CACGACGCTGCGAGCAGCGGTCTGACCGTGTCCGAAACGAGAGCTCCCAAGAAATGCTGCTG 8509  
Qy 901 AAACGCGGCTGCGGAGACGAGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
Db 8510 AAACGCGGCTGCGGAGACGAGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8569  
Qy 961 GGCACGGGCGACGAGCGCTTGTGACCCCATCGAAATCCAAAGCTGGAATGCGGTATACGGC 1020  
Db 8570 GGCACGGGCGACGAGCGCTTGTGACCCCATCGAAATCCAAAGCTGGAATGCGGTATACGGC 8629  
Qy 1021 CTGCGGCGAGATGTCGCCACCGCCGCTGCTGATTCGGGTGCTGAGAACCAACTTGGCCAT 1080  
Db 8630 CTGCGGCGAGATGTCGCCACCGCCGCTGCTGATTCGGGTGCTGAGAACCAACTTGGCCAT 8689  
Qy 1081 CTTGAGTATGCTGCGGAGATCACTGAGGCTGCTGAAAGGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
Db 8690 CTTGAGTATGCTGCGGAGATCACTGAGGCTGCTGAAAGGCTGCTGCTGCTGCTGCTGCTGCTG 8749  
Qy 1141 CAGATTCTGCGCACTTCCACGCGCAGCGCTGCAACCCCGGATTCATGAGGCTGATCTT 1200  
Db 8750 CAGATTCTGCGCACTTCCACGCGCAGCGCTGCAACCCCGGATTCATGAGGCTGATCTT 8809  
Qy 1201 CGGCTGACCTGACGCGCGCCCGGACACCTGCGCTGCAATGCGCCGCGACGGCGC 1260  
Db 8810 CGGCTGACCTGACGCGCGCCCGGACACCTGCGCTGCAATGCGCCGCGACGGCGC 8869  
Qy 1261 GGGGTGACCTGCTGCGATGAGGCGGACCAACGCGCAGGTGCTGGAAGAGGCGCGC 1320  
Db 8870 GGGGTGACCTGCTGCGATGAGGCGGACCAACGCGCAGGTGCTGGAAGAGGCGCGC 8929  
Qy 1321 GCGGCGAGTGCACACCGCGCGCGCGGAGGACCGGCGAGAGCTGCTGCTGCGCA 1380  
Db 8930 GCGGCGAGTGCACACCGCGCGCGCGGAGGACCGGCGAGAGCTGCTGCTGCGCA 8989  
Qy 1381 AGAACCGGTGACGCTTGTGATGACAGCGCGCGCGGCTGCGGACATCTGGAACCTTAC 1440  
Db 8990 AGAACCGGTGACGCTTGTGATGACAGCGCGCGCGGCTGCGGACATCTGGAACCTTAC 9049  
Qy 1441 CTTTGCAGTGTCTGCGGAGATGAGGCTTCACTGCTGCGAGACGACGCGCGCGATGAG 1500  
Db 9050 CTTTGCAGTGTCTGCGGAGATGAGGCTTCACTGCTGCGAGACGACGCGCGCGATGAG 9109  
Qy 1501 CACCGGCTGCGGCTGCGCGACGCTGAGGAGAGGCTGCGGCGACGCTTGGACGCTGCG 1560  
Db 9110 CACCGGCTGCGGCTGCGCGACGCTGAGGAGAGGCTGCGGCGACGCTTGGACGCTGCG 9169  
Qy 1561 GCGCAGGAGACAGAGCTGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1620  
Db 9170 GCGCAGGAGACAGAGCTTCCCGGCTGCGGCTGCGGCTGCGGCTTCTTCCACGCGGCAAG 9229  
Qy 1621 CTGCGCTTCTTCTTACCGGACAGGCGGCGAGACGCTGCGGAGATGCGCGCTGCTGAC 1680  
Db 9230 CTGCGCTTCTTCTTACCGGACAGGCGGCGAGACGCTGCGGAGATGCGCGCTGCTGAC 9289  
Qy 1681 GATGTATGCTCGGCTTCCGCGAGGCTTTCGACCTGTGCGTGAAGCTTTCACACGAGAG 1740

Db	9290	GATGTATGTCCTCCGGCTTCGCCAGAGCGCTTGACACTGTGGGTAGAGGCTGTTCACACAGAGG	9349
Qy	1741	CTGCAGCCGGCCGCTCGCGAGAGTGATGTGGAGCGAACCAGCCAGCGTCGACCGCCGCTG	1800
Db	9350	CTGCACCGGCGCGCTCCGCGAGGTGATGTGGAGCGAACCAGCCAGCGTCGACCGCCGCTG	9409
Qy	1801	CTGCACCAAGACGCTTCACCCAGCCGGCGCTGTTCACCTTCGAAATATGCGCTCGCCGCG	1860
Db	9410	CTGCACCAAGACGCTTCACCCAGCCGGCGCTGTTCACCTTCGAAATATGCGCTCGCCGCG	9469
Qy	1861	CTGTGGCGGTCGTGTGGGGGTATAGAGCCGGAAGTGTGCGCGGCAATATGACCTGGTGAAGTG	1920
Db	9470	CTGTGGCGGTCGTGTGGGGGTATAGAGCCGGAAGTGTGCGCGGCAATATGACCTGGTGAAGTG	9529
Qy	1921	GTGGCTCGCTCGGTGAGCGGCGGTGTTCCTCGCTTGAGAGACGCGGTGTTCCTGTGAGCTGCG	1980
Db	9530	GTGGCTCGCTCGGTGAGCGGCGGTGTTCCTCGCTTGAGAGACCGCGGTGTTCCTGTGAGCTGCG	9589
Qy	1981	CGCGGGGCGCTGATGACAGGCGCTGCGCGCGCGCGCGGCGATGTGTTCATACAGAGCGCGG	2040
Db	9590	CGCGGGGCGCTGATGACAGGCGCTGCGCGCGCGCGCGGCGATGTGTTCATACAGAGCGCGG	9649
Qy	2041	GAGGCGCATGTGGGTGTGGGGTGGCGGCGGACGACGCTGGTGTTCATACCGCGCGCTG	2100
Db	9650	GAGGCGCATGTGGGTGTGGGGTGGCGGCGGACGACGCTGGTGTTCATACCGCGCGCTG	9709
Qy	2101	AACGCTCCGGACACAGGTGATCATTCGCGGAGCGCGGCGCAACCCTGTGATCGATTCGCGCG	2160
Db	9710	AACGCTCCGGACACAGGTGATCATTCGCGGAGCGCGGCGCAACCCTGTGATCGATTCGCGCG	9769
Qy	2161	GCGATGACCCGCGCGGGGCGCGAACCAGGCGCTCCACGTCTCGATGCGTTCCACTCA	2220
Db	9770	GCGATGACCCGCGCGGGGCGCGAACCAGGCGCTCCACGTCTCGATGCGTTCCACTCA	9829
Qy	2221	CCGCTCATGGCCCCGGAATGCTGAGAGCGCTTGGGCGGTGTGGCCGAGTCGCTGAGCTACCGG	2280
Db	9830	CCGCTCATGGCCCCGGAATGCTGAGAGCGCTTGGGCGGTGTGGCCGAGTCGCTGAGCTACCGG	9889
Qy	2281	CGGCGGTGATTCGTCCTGTGTCAGCAATCTGAGCGGGAAGGCTTGACACAGACGAGTGAAC	2340
Db	9890	CGGCGGTGATTCGTCCTGTGTCAGCAATCTGAGCGGGAAGGCTTGACACAGACGAGTGAAC	9949
Qy	2341	TCGCGGAGGCTATTGGGTGCGCAACGCGCGAGAGGTGGTGGCTTCGCGAGTGAAGTGAAG	2400
Db	9950	TCGCGGAGGCTATTGGGTGCGCAACGCGCGAGAGGTGGTGGCTTCGCGAGTGAAGTGAAG	10009
Qy	2401	GCGCTGACACGCGCGGTCGCGGACCTTGTGTGAAGTCCGTTCGAAATGACGCTGCTC	2460
Db	10010	GCGCTGACACGCGCGGTCGCGGACCTTGTGTGAAGTCCGTTCGAAATGACGCTGCTC	10069
Qy	2461	GCGCTGAGGCTGCTGTCATGACGCGGAAGCGCCGGCGCGCTGCTGCATATGTCGCGCGCT	2520
Db	10070	GCGCTGAGGCTGCTGTCATGACGCGGAAGCGCCGGCGCGCTGCTGCATATGTCGCGCGCT	10129
Qy	2521	GGGCGGTGACGAGCGCGGACCGGTGCTGAGAGCGCTCGGCGGTCGTGGGCGCGTCGATGGG	2580
Db	10130	GGGCGGTGACGAGCGCGGACCGGTGCTGAGAGCGCTCGGCGGTCGTGGGCGCGTCGATGGG	10189
Qy	2581	CTGGTCTCTCTGGGCGCGGCTCTTTCCTCAGGAGGCGCGCGGTCGCTCCACGTAAC	2640
Db	10190	CTGGTCTCTCTGGGCGCGGCTCTTTCCTCAGGAGGCGCGCGGTCGCTCCACGTAAC	10249
Qy	2641	CGTTGGACAGGCGGACGCTACTGGATGTGACACGAAAGCGGACGACGCGGCGCGTGGCGAC	2700
Db	10250	CGTTGGACAGGCGGACGCTACTGGATGTGACACGAAAGCGGACGACGCGGCGCGTGGCGAC	10309
Qy	2701	CGCGGTGCTCTCGGAGCGGGTTCACGACGAGTTCGAGAGAGGAGGCGCGGTCGCGCGGCG	2760
Db	10310	CGCGGTGCTCTCGGAGCGGGTTCACGACGAGTTCGAGAGAGGAGGCGCGGTCGCGCGGCG	10369
Qy	2761	GACCGGCGACGCGTCCGCTTCGACATTCGCGCGCTCCGAGGCGGACGCGCGGAGAAAGTTC	2820
Db	10370	GACCGGCGACGCGTCCGCTTCGACATTCGCGCGCTCCGAGGCGGACGCGCGGAGAAAGTTC	10429

QY	2821	GAGCGCGCGGCGACACCGTCCGTTCCGGCTCCAGATTCGAGTAGACGACGCTGCTTATATAC	2880
Db	10430	GAGCGCGCGGCGACCGTCCGTTCCGGCTCCAGATTCGAGTAGACGACGCTGCTTATATAC	10489
QY	2881	CTCGTGTTCGGGATCAGCGAGCGCGCGCCCTGCTCTGAGCGAGGTCAGATTCGCGCTC	2940
Db	10490	CTCGTGTTCGGGATCAGCGAGCGCGCGCCCTGCTCTGAGCGAGGTCAGATTCGCGCTC	10549
QY	2941	GACCGCGCGGCGCTCAGCTTCAATGATGTCAGCTTCGCTGGGCAATGTTGCCGACGAC	3000
Db	10550	GACCGCGCGGCGCTCAGCTTCAATGATGTCAGCTTCGCTGGGCAATGTTGCCGACGAC	10609
QY	3001	CTGCGCGGAAAGCCAAACCTCCGCTGCTGTCGAGGGGAGTGCGCGCGGCGCATTCGTC	3060
Db	10610	CTGCGCGGAAAGCCAAACCTCCGCTGCTGTCGAGGGGAGTGCGCGCGGCGCATTCGTC	10669
QY	3061	GCCGTGGCGAGGGCGCTGAACGAGCTCTGTGTGGGCAACCGGATCATCGCCCTTCGGCG	3120
Db	10670	GCCGTGGCGAGGGCGCTGAACGAGCTCTGTGTGGGCAACCGGATCATCGCCCTTCGGCG	10729
QY	3121	GAGAGCTTTGCTACCCACGTCACCAAGTGGGCTGCGCTGCTGTGCTCGGCTCAAGCG	3180
Db	10730	GAGAGCTTTGCTACCCACGTCACCAAGTGGGCTGCGCTGCTGTGCTCGGCTCAAGCG	10789
QY	3181	CTCTCGCGCATTCAGAGCGGCGCGCATGCGCTGCGCTACCTGACGCGATGATACGCGCTC	3240
Db	10790	CTCTCGCGCATTCAGAGCGGCGCGCATGCGCTGCGCTACCTGACGCGATGATACGCGCTC	10849
QY	3241	GACAGAAATAGCCCGCTTACAGCTGGGGGAGCGGGTGTCTATCATATGCGCGACCGGCGGG	3300
Db	10850	GACAGAAATAGCCCGCTTACAGCTGGGGGAGCGGGTGTCTATCATATGCGCGACCGGCGGG	10909
QY	3301	GTCGAGTCTCGCGCGGCGAGTGGGGGCGACAGAGTGGAGCGAGGTCATCGACGACGCGCC	3360
Db	10910	GTCGAGTCTCGCGCGGCGAGTGGGGGCGACAGAGTGGAGCGAGGTCATCGACGACGCGCC	10969
QY	3361	GGCACGCGCCGAGAAACGCGGCTTACCTGAGTCGCTGGGCGTGGCGATGTGAGCGCATTC	3420
Db	10970	GGCACGCGCCGAGAAACGCGGCTTACCTGAGTCGCTGGGCGTGGCGATGTGAGCGCATTC	11029
QY	3421	CGCTCCGGAACCGGTTCTGTCGCGCGAGCGCGCGCTGGACGCGGGCGAGGGAATTAACGCTC	3480
Db	11030	CGCTCCGGAACCGGTTCTGTCGCGCGAGCGCGCGCTGGACGCGGGCGAGGGAATTAACGCTC	11089
QY	3481	GTCGCTCAACTCGGCTCTCGGCGAGCTGATTCGACAGAGTTTCAATCTCCTGGATTCGAC	3540
Db	11090	GTCGCTCAACTCGGCTCTCGGCGAGCTGATTCGACAGAGTTTCAATCTCCTGGATTCGAC	11149
QY	3541	GCGCGGTTTGTGAGAGCTCGGCAAGCGCGACTGTTACGCGGATPACCAAGTTCGGGCTCGG	3600
Db	11150	GCGCGGTTTGTGAGAGCTCGGCAAGCGCGACTGTTACGCGGATPACCAAGTTCGGGCTCGG	11209
QY	3601	CCGTTTCTGTGCAATCTCTCTCTGCGTGGTGAATCTTCGCGGGGATGATGCTCGAGCGG	3660
Db	11210	CCGTTTCTGTGCAATCTCTCTCTGCGTGGTGAATCTTCGCGGGGATGATGCTCGAGCGG	11269
QY	3661	CCGCGCGGAGTCCGTGCGCTCTTGAAGAGCTTCTCGGCTGATCGCGGACGCGGTTC	3720
Db	11270	CCGCGCGGAGTCCGTGCGCTCTTGAAGAGCTTCTCGGCTGATCGCGGACGCGGTTC	11329
QY	3721	ACCCGCTCCGCGCATTCGCGAGCGCTCCCGCATGCGCGCGGATGCGGTCGTTCCGGAGCATG	3780
Db	11330	ACCCGCTCCGCGCATTCGCGAGCGCTCCCGCATGCGCGCGGATGCGGTCGTTCCGGAGCATG	11389
QY	3781	GCGCAGGCGCAGCATCTTGGGAGAGCTCGTACTCAGCGTGGGTGACCCGAGGTCAGATTC	3840
Db	11390	GCGCAGGCGCAGCATCTTGGGAGAGCTCGTACTCAGCGTGGGTGACCCGAGGTCAGATTC	11449
QY	3841	CGTATTTCCAAACCAACGACGAGCGCGCGCTCAGCGGGGATCGGGGACTGCTCGACAGG	3900
Db	11450	CGTATTTCCAAACCAACGACGAGCGCGCGCTCAGCGGGGATCGGGGACTGCTCGACAGG	11509

[illegible]

protein\_id="AAFG2878.1"  
/db\_xref="GI:7453555"  
/translation="VRGHVEAFADPGVPRVLYDNLKMAVVERHGDIAFHPTLAL  
SADYFERPRPAVAFGRNEKGVREPAITYREBGFPEBAVAYADLDGRKQATERNISSAL  
DRSWDEARVAFVQAFDERSVLIRHPTDFPDEREFAVVGKTPVAPFLNLSVPH  
DTRRTTIVLADLSQVRIADNGQIVAVHVSVMGQOIEQEPHEQIVYDKREARERH  
GLDLAPARASSQAFILRIVAERGDNVSAIARLLQGLDAGAELEBALVEAVERDPT  
HIGAVROVIDRREBERHLPVPSIVPTVRGHAAVLVTPHSLTTYDALKKQPTP"  
989..1501  
/note="orfB"  
/codon\_start=1  
/transl\_table=1  
/product="putative transposase"  
/protein\_id="AAFG2879.1"  
/db\_xref="GI:7453556"  
/translation="MTDILPTETDKLKSGLFGLLACMGQLADKPMPLREVLAIERE  
RHKSLERLRKNSRVAAFKPMTDSDSMPKKIDREAVDDIYDSRYADLLEVVTRRD  
AOKRLISTNKAFAFDMQGVFPFAACVTVLDRLVHRAVEIEIEASRYLKEAKELAT  
RKORRTKKH"  
1998..6263  
/gene="epoB"  
/gene="epoB"  
1998..6263  
/note="type I polyketide synthase; epoH1lone PKS loading  
module"  
/codon\_start=1  
/transl\_table=1  
/product="EpoB"  
/protein\_id="AAFG2880.1"  
/db\_xref="GI:7453557"  
/translation="MADRPTERAEDPIATVAGCRLPGGVLDLSGFMTLLDSRPDV  
GVYAEKRDAAAFDPDLDAKGTPTPRASFLDVGCFDPSFGISPRELRMDPMHR  
LLLECHMALNNALIASALVGTETGVFTGICSEYERALPRATSAEIDHAGSLGTM  
PSVAGARISYVLGRPCVADVTAIVSSILVAHLACQSLSGECSTALACGVLMISF  
STLWLRSTRLATDGRCKAFSAEDGFGRGEGCAVVLRKLGARADGRIILAVIG  
SAIHNDGASCLATVPNGSSOEIVLKRLADAGCASSGVYEAHGTTGADTLEIOA  
LNAVYGRDVAFTLLIGSVKTNLGHBEVASGITGLKVLISLQHQIIPHLHAOLIN  
PRIKGDRLRTVTARPTPMDPTPRAGVSSRMSGTNAHVLEAPATCTPPAPE  
RPAELVLSATTAALDAHAHARLDHLETPSQCLDVAFSLATTSAMHRLAVAT  
SSEGRILADAAAGQTPGPGVGRGIAOSSSGKLAFLPTGGQAOTLGMGRULVYVMP  
REAPDLCTRLFQELDRPLREVMWAEPAEDDALDCTATOPALPFEELALMAWS  
MGVPELVLAGHSIGELVAACVAGVSLIEDVFLVAAGLMTQAPGAGVAAIAAPA  
DVAAVLAAPHASVSIIAVNGDDOVIAGAOOPVAAIAAAGAARAKALHVSIAFPA  
PLMAPLEAPERVAESYRPSIVTIVNSLGRKGVDESPGVGMHAEEVVRPAG  
VKALHAAGAVFSEVGRKSTLGLVPCREDPAPALLASSRGRDEPATVLEAGLGM  
AVGLVSAVAGFLPVGGRVPLPTYPMQREYRIDTKADDARAGURRPAAGHBEVEG  
GAVNGGRSARLDHPPESGRREKVEAAGDRPRLEIDEPVGLDRVLVLETERAAG  
LGEVETADNAGLGFNDVOLAGVPPDLPGKPNPPLLGEGCGRIVAAEGEONGLV  
GEOVILIASCAAFATVHTTSAALVLRPOLASATEAAMPVAYLTAWYALDGLARLP  
GERVLIHAATGCGVGLAIVQAKHGAHVHATAGTPEKRAVLESGRVYSDSSGDRP  
ADVPAWTEGEGVDVNLISLSELIDKSFNLIRSHGRVEVLEKGDICADNOLGRLPPLR  
NLSFLVDLRMMLERPARVAFLEELIGLIAAGVFPPIATLPIARVADARSMQO  
AOKHGKVLTLTGDEPVOIRIPTFHAGAGSTGDRDLDRILASAAPAAKAAALAEFLRQ  
VSOVLARPEITVGALEALFTRLGMSLMAVEENRRIEASLKLKLTSTFLTSPTNALLT  
ONLIDALATALSERVAENLRAGVQSDPFSAGADQWEITAI"  
6260..10492  
/gene="epoB"  
/gene="epoB"  
6260..10492  
/note="non ribosomal peptide synthetase; epoH1lone PKS  
module 1"  
/codon\_start=1  
/transl\_table=1  
/product="EpoB"  
/protein\_id="AAFG2881.1"  
/db\_xref="GI:7453558"  
/translation="MTINQLNLNEHGGVLAADGERLQIQAPRNALNPULAIRISH  
KSTILTMLRQLPAESIIVPAPAEHRHVPPLTIDQGSWLRTGAFVPSITAHARETD  
CTDIDVAALAFKRRVPAHDMLEAAHLPMMQOITEKVAADIETIDRLGDSSTREA  
RLVSLRDMSSRIIVYTERPPLVAVVAARLDEOQRTVLSIDLINVDIGSLISIFKMDL  
SFYDEPESLIVLELSDRYDYLAEKSKSEAHORSMDYMRVAALEAPPMPLMKAD  
PSTLAREIFRRTBEWLPBDSWSRLKQVGRGRLTPGVILAAESEVIGRWASAPRFL  
NTTLFNRLVHPRVANDITGDFTSNVLDITDTRDKSEQAQAKIQEOLWAMMHCVS



QY 841 CACGACGCGTGAAGAGCGGCTTCTGACCCGCGGAAACGGGAGCTCCCAAGAAATCGTCGCG 900  
| | | | |  
Db 2838 CACGACGAGCGAGAGCGGCTCTGACCTGCGAAACGGGAGCTCCCAAGAAATCGTCGCG 2897  
QY 901 AAACGGGCGCTTGGCGGAGACGACGCTGCGCGCGCTCTTTCGGTGGTTATGTCGAGGACAC 960  
| | | | |  
Db 2898 AAACGGGCGCTTGGCGGAGACGACGCTGCGCGCGCTCTTTCGGTGGTTATGTCGAGGACAC 2957  
QY 961 GGCACGGGACGACGCTTGGTGAACCCCATCGAAATCGAAAGCTCTGAATGCGGTATACGGC 1020  
| | | | |  
Db 2958 GGCACGGGACGACGCTTGGTGAACCCCATCGAAATCGAAAGCTCTGAATGCGGTATACGGC 3017  
QY 1021 CTCGGGCGGAGATGTCGACGCGCGCTGCTGATCGGCTCGGTGAACCAACCTTGCGCAT 1080  
| | | | |  
Db 3018 CTCGGGCGGAGATGTCGACGCGCGCTGCTGATCGGCTCGGTGAACCAACCTTGCGCAT 3077  
QY 1081 CTTGAGTATGCTCGGGGATCACTGGGCTGCTGAAGTGTCTTGTCTCTTGACGACGGG 1140  
| | | | |  
Db 3078 CTTGAGTATGCTCGGGGATCACTGGGCTGCTGAAGTGTCTTGTCTCTTGACGACGGG 3137  
QY 1141 CAGATTCCTGCGCACCTTCCACGCGACGCGGCTGAAACCCCGGATCTCATGSGGTATCTT 1200  
| | | | |  
Db 3138 CAGATTCCTGCGCACCTTCCACGCGACGCGGCTGAAACCCCGGATCTCATGSGGTATCTT 3197  
QY 1201 CGGCTGACCTGTCACGCGCGCGCGGACACCTGTCGCGGACCTGAATACCGCGGACGCGCG 1260  
| | | | |  
Db 3198 CGGCTGACCTGTCACGCGCGCGCGGACACCTGTCGCGGACCTGAATACCGCGGACGCGCG 3257  
QY 1261 GGGGCTGAGCTCTTTCGCGCATGACGCGGACCAACGCGCATGTCGTCGTAAGAGGCGCG 1320  
| | | | |  
Db 3258 GGGGCTGAGCTCTTTCGCGCATGACGCGGACCAACGCGCATGTCGTCGTAAGAGGCGCG 3317  
QY 1321 GCGGCGAGCTGTCACACCGCGCGCGCGCGGACCGGACGACGCTGTCGTCGTCGTCGTC 1380  
| | | | |  
Db 3318 GCGGCGAGCTGTCACACCGCGCGCGCGGACCGGACGACGCTGTCGTCGTCGTCGTCGTC 3377  
QY 1381 AGGACCGGCTGACCTTGGATGACACAGGCGGCGCGGCTGCGGACCATCTGAGAGCTTAC 1440  
| | | | |  
Db 3378 AGGACCGGCTGACCTTGGATGACACAGGCGGCGCGGCTGCGGACCATCTGAGAGCTTAC 3437  
QY 1441 CTTTCGACGATGTCGTCGCGATGTCGCTTCACTTCGCGACGACGCGGACGCGATGAG 1500  
| | | | |  
Db 3438 CTTTCGACGATGTCGTCGCGATGTCGCTTCACTTCGCGACGACGCGGACGCGATGAG 3497  
QY 1501 CACCGGCTGCGCGTGGCGGCGACGTCGAGAGGAGGCGCTGCGGACGCTTGAACGCTGCG 1560  
| | | | |  
Db 3498 CACCGGCTGCGCGTGGCGGCGACGTCGAGAGGAGGCGCTGCGGACGCTTGAACGCTGCG 3557  
QY 1561 GCGGAGGAGACAGACGTCGCGCGTGGCGGCGGTCGAGATGCGGATTCCTCAACGCGGACAG 1620  
| | | | |  
Db 3558 GCGGAGGAGACAGACGTCGCGCGTGGCGGCGGTCGAGATGCGGATTCCTCAACGCGGACAG 3617  
QY 1621 CTCGCTTCTTCTTCAACGCGACAGGAGGAGCGACAGCTGAGGACGCGTGGCGCTGTAC 1680  
| | | | |  
Db 3618 CTCGCTTCTTCTTCAACGCGACAGGAGGAGCGACAGCTGAGGACGCGTGGCGCTGTAC 3677  
QY 1681 GATGATGTCGTCGCTTTCGCGAGGCGCTTTCGACCTGTCGTCGTCGTCGTCGTCGTCGTC 1740  
| | | | |  
Db 3678 GATGATGTCGTCGCTTTCGCGAGGCGCTTTCGACCTGTCGTCGTCGTCGTCGTCGTCGTC 3737  
QY 1741 CTCGACGCGGCGCTCGCGAGGTCGATGTCGCGGCGGAAACGCGGACGAGTCGACGCGCGCTG 1800  
| | | | |  
Db 3738 CTCGACGCGGCGCTCGCGAGGTCGATGTCGCGGCGGAAACGCGGACGAGTCGACGCGCGCTG 3797  
QY 1801 CTCGACGAGACAGCTTTCACCCAGCGCGGCTGTTTCACCTTCGAAATGTCGCTCGCGCG 1860  
| | | | |  
Db 3798 CTCGACGAGACAGCTTTCACCCAGCGCGGCTGTTTCACCTTCGAAATGTCGCTCGCGCG 3857  
QY 1861 CTGTGCGGCTGTCGTGGGTGTAGAGCGGAGTCGTCGCGGACATGACATGTCGTCGTCGTCG 1920  
| | | | |  
Db 3858 CTGTGCGGCTGTCGTGGGTGTAGAGCGGAGTCGTCGCGGACATGACATGTCGTCGTCGTCG 3917  
QY 1921 GTGCGTGCCTGCTGTCGTCGCGGCGCTGTTCTTCGCTTGAAGACGCGGTGTTCTCTGTCGCTGCG 1980  
| | | | |

Db 3918 GTGCTGCTGCTGTCGTCGCGGCTGTTCTTCGCTTGAAGACGCGGTCTCTGTCGTCGCG 3977  
| | | | |  
QY 1981 CGCGGCGCTGATGACAGCGCTGCGGCGCGGCGGAGTCGATGTCGATCGAGCGCGCG 2040  
| | | | |  
Db 3978 CGCGGCGCTGATGACAGCGCTGCGGCGCGGCGGAGTCGATGTCGATCGCGCGCGCG 4037  
QY 2041 GAGGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2100  
| | | | |  
Db 4038 GAGGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4097  
QY 2101 AACGCTCGGACAGAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2160  
| | | | |  
Db 4098 AACGCTCGGACAGAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4157  
QY 2161 GCGATGCGCGCGCGCGGCGGAGCAAGAGCGCTTCCAGTCGTCGATGCTTCCACTCA 2220  
| | | | |  
Db 4158 GCGATGCGCGCGCGCGGCGGAGCAAGAGCGCTTCCAGTCGTCGATGCTTCCACTCA 4217  
QY 2221 CGGCTCATGCGCGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2280  
| | | | |  
Db 4218 CGGCTCATGCGCGCGGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4277  
QY 2281 CGGCGCTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2340  
| | | | |  
Db 4278 CGGCGCTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4337  
QY 2341 TCGCGGCGCTTATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2400  
| | | | |  
Db 4338 TCGCGGCGCTTATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4397  
QY 2401 GCGCTGACAGCGCGCGGCGGCGGACCTTTCGTCGAGTCGTCGTCGTCGTCGTCGTCGTCGTC 2460  
| | | | |  
Db 4398 GCGCTGACAGCGCGCGGCGGCGGACCTTTCGTCGAGTCGTCGTCGTCGTCGTCGTCGTCGTC 4457  
QY 2461 GCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2520  
| | | | |  
Db 4458 GCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4517  
QY 2521 GCGGCTGACAGCGCGGCGGACCTTTCGTCGAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2580  
| | | | |  
Db 4518 GCGGCTGACAGCGCGGCGGACCTTTCGTCGAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4577  
QY 2581 CTGTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2640  
| | | | |  
Db 4578 CTGTCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4637  
QY 2641 CTTTCGACGAGCGCTTACTGATGTCGACAGAAAGCGGACGACGCGCGCGCTGTCGAC 2700  
| | | | |  
Db 4638 CTTTCGACGAGCGCTTACTGATGTCGACAGAAAGCGGACGACGCGCGCGCTGTCGAC 4697  
QY 2701 GCGCGTCTTCGCGAGCGGTCGACGAGAGTCGAGAGGAGGCGCGCTGTCGCGCGCGC 2760  
| | | | |  
Db 4698 GCGCGTCTTCGCGAGCGGTCGACGAGAGTCGAGAGGAGGCGCGCTGTCGCGCGCGC 4757  
QY 2761 GACCGGCGGAGCGCTCGGCTGACGATCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 2820  
| | | | |  
Db 4758 GACCGGCGGAGCGCTCGGCTGACGATCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 4817  
QY 4818 GAGGCGGCGGAGCGCTCGGCTGACGATCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 4877  
| | | | |  
Db 4878 CTGTGCTTTCGAGTCACGAGAGCGGCGCGCTTTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTC 4937  
QY 2941 GACCGGCGGAGCGCTTCGATGATGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTC 3000  
| | | | |  
Db 4938 GACCGGCGGAGCGCTTCGATGATGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTCGAGTC 4997  
| | | | |  
QY 3001 CTGCGGAGAAAGCCCAACCTTCGCTGCTGTCGAGGCGAGTCGCGCGGCGCATGCTC 3060  
| | | | |



Dh 4998 CTCGCGGAAAACCCAACTCTCGCTGCTCTGAGGCGAGTGGCCGCGGCGCATGTC 5057  
Qy 3061 GCCCTGGGCGAGGCGCTGAAAGCGCTCTGAGTGGGCGCAACCGGTCATGCGCTTTCGCG 3120  
Dh 5058 GCGGTGGGCGAGGCGCTGAAAGCGCTCTGAGTGGGCGCAACCGGTCATGCGCTTTCGCG 5117  
Qy 3121 GAGGCTTTGCTACCAAGTCAACAGTGGGCGCTGAGTGGGCGCTGAGTGGGCGCTGAG 3180  
Dh 5118 GAGGCTTTGCTACCAAGTCAACAGTGGGCGCTGAGTGGGCGCTGAGTGGGCGCTGAG 5177  
Qy 3181 CTCCTGGGCGAGTGGGCGGCGCGCATGCGCTGAGTGGGCGCTGAGTGGGCGCTGAG 3240  
Dh 5178 CTCCTGGGCGAGTGGGCGGCGCGCATGCGCTGAGTGGGCGCTGAGTGGGCGCTGAG 5237  
Qy 3241 GACGAAATAGCCGCTTTCAGCGCGGCGGAGCGGCTGATCTCATGCGCGCGCGCGG 3300  
Dh 5238 GACGAAATAGCCGCTTTCAGCGCGGCGGAGCGGCTGATCTCATGCGCGCGCGCGG 5297  
Qy 3301 GTCGCTCTCGCGCGGCTGAGTGGGCGCGAGCGTGGGAGCGGAGTCTCATGCGCGCG 3360  
Dh 5298 GTCGCTCTCGCGCGGCTGAGTGGGCGCGAGCGTGGGAGCGGAGTCTCATGCGCGCG 5357  
Qy 3361 GGCACGCGCGCGAGAAACCGCGCTTACCTGAGTGGGCGCTGAGTGGGCGCTGAG 3420  
Dh 5358 GGCACGCGCGCGAGAAACCGCGCTTACCTGAGTGGGCGCTGAGTGGGCGCTGAG 5417  
Qy 3421 CGCTCGGACCGGCTCTGCGCGCGAGCGTGGGCGCGAGCGGCGGAGTGGAGTGC 3480  
Dh 5418 CGCTCGGACCGGCTCTGCGCGCGAGCGTGGGCGCGAGCGGCGGAGTGGAGTGC 5477  
Qy 3481 GTGCTCAACTCGCTCTCGGCGGAGCTGATCGAACAAGTTCATCTCTGCGATCGAC 3540  
Dh 5478 GTGCTCAACTCGCTCTCGGCGGAGCTGATCGAACAAGTTCATCTCTGCGATCGAC 5537  
Qy 3541 GCGCGGCTTTGAGAGCTCTGCGGAGCGGAGCTGATCGGAGTAAACAGCTTGGGCTCG 3600  
Dh 5538 GCGCGGCTTTGAGAGCTCTGCGGAGCGGAGCTGATCGGAGTAAACAGCTTGGGCTCG 5597  
Qy 3601 CCGTCTCTGCGCAATCTCTCTCTCTGCTGAGTCTGCGGCGGAGTAAACAGCTTGGGCT 3660  
Dh 5598 CCGTCTCTGCGCAATCTCTCTCTCTGCTGAGTCTGCGGCGGAGTAAACAGCTTGGGCT 5657  
Qy 3661 CCGGCGGCGGCTCTGCGGCTCTTGGAGAGTCTCTGCGGCTGATCGGCGGAGTCTG 3720  
Dh 5658 CCGGCGGCGGCTCTGCGGCTCTTGGAGAGTCTCTGCGGCTGATCGGCGGAGTCTG 5717  
Qy 3721 ACCCTTCCCCCATCTGCGAGCTCTGCGGCTCTGCGGCTGATCGGCGGAGTCTG 3780  
Dh 5718 ACCCTTCCCCCATCTGCGAGCTCTGCGGCTCTGCGGCTGATCGGCGGAGTCTG 5777  
Qy 3781 GCGGAGGCGAGCATCTTGGGAGCTGTAACAAGTGGGTAACCGGAGGTCAGATC 3840  
Dh 5778 GCGGAGGCGAGCATCTTGGGAGCTGTAACAAGTGGGTAACCGGAGGTCAGATC 5837  
Qy 3841 CGATTTCCAAACCAACGAGGCGCGGCGGCTCTGCAACCGGAGTCTGCGGAGTCTG 3900  
Dh 5838 CGATTTCCAAACCAACGAGGCGCGGCGGCTCTGCAACCGGAGTCTGCGGAGTCTG 5897  
Qy 3901 CTGCGCTGAGTCTGCGGCGCGGCGGCGGCTCTGCAACCGGAGTCTGCGGAGTCTG 3960  
Dh 5898 CTGCGCTGAGTCTGCGGCGCGGCGGCGGCTCTGCAACCGGAGTCTGCGGAGTCTG 5957  
Qy 3961 GTCCTGAGGCTGCGGCGCGGCGGCGGCGGCTCTGCAACCGGAGTCTGCGGAGTCTG 4020  
Dh 5958 GTCCTGAGGCTGCGGCGCGGCGGCGGCGGCTCTGCAACCGGAGTCTGCGGAGTCTG 6017  
Qy 4021 CTCGCGATGAGCTGCTCATGCGCGTGGAGTGGCAATCTATCGAGCGAGCGCTCAG 4080  
Dh 6018 CTCGCGATGAGCTGCTCATGCGCGTGGAGTGGCAATCTATCGAGCGAGCGCTCAG 6077  
Qy 4081 CTGAAAGTCTGAGAGAGTCTCTGCTCAAGTCTGCGGCGGAGTCTGCTGTTGGCCAAAC 4140  
Dh 6078 CTGAAAGTCTGAGAGAGTCTCTGCTCAAGTCTGCGGCGGAGTCTGCTGTTGGCCAAAC 6137

Qy 4141 CTGTTGATGCTCTGCGCAGAGCTCTCTCTTGGAGCGGCTGGCGCGGAACTTACGG 4200  
Dh 6138 CTGTTGATGCTCTGCGCAGAGCTCTCTCTTGGAGCGGCTGGCGCGGAACTTACGG 6197  
Qy 4201 GCAAGCGCTGAAACGACTTCTGCTTCACTGGGCGCGAGATCAAGACTGGAAATATTGCC 4260  
Dh 6198 GCAAGCGCTGAAACGACTTCTGCTTCACTGGGCGCGAGATCAAGACTGGAAATATTGCC 6257  
Qy 4261 CTATGA 4266  
Dh 6258 CTATGA 6263

RESULT 9  
AR172664  
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001  
DEFINITION Sequence 2 from patent US 6303342.  
ACCESSION AR172664  
VERSION AR172664.1 GI:17912155  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 71989)  
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.  
TITLE Recombinant methods and materials for producing epothenones C and D  
JOURNAL Patent: US 6303342-A 2 16-OCR-2001;  
FEATURES  
source 1. 71989  
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others  
ORIGIN

Query Match 98.2%; Score 4187.6; DB 6; Length 71989;  
Best Local Similarity 98.9%; Pred. No. 0;  
Matches 4217; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 GTGCGGATGCTGCTCCATCGAGCGCGAGCGGAGTCCGATGCGATGCTGAGCGAGT 60  
Dh 1998 GTGCGGATGCTGCTCCATCGAGCGCGAGCGGAGTCCGATGCGATGCTGAGCGAGT 2057  
Qy 61 TGCCTGTGCGCGGCTGATGATGATGAGCGGCTTCTGAGCGCTCTGAGGCGCTCG 120  
Dh 2058 TGCCTGTGCGCGGCTGATGATGATGAGCGGCTTCTGAGCGCTCTGAGGCGCTCG 2117  
Qy 121 CGGACACCGTGGGCGAGTCCCGCGGAGCGCTGGATCAGAGCGTGGTGTATCCC 180  
Dh 2118 CGGACACCGTGGGCGAGTCCCGCGGAGCGCTGGATCAGAGCGTGGTGTATCCC 2177  
Qy 181 GACCCGATGCGCGGCGGAGAGCGCGGTTACGCGGCGATCTTCTGAGCGAGTAGCC 240  
Dh 2178 GACCCGATGCGCGGCGGAGAGCGCGGTTACGCGGCGATCTTCTGAGCGAGTAGCC 2237  
Qy 241 TGCCTGACCGCTCTTCTTCTGCGATCTGCGCTGCGAGCGGCTGCGAGTACCTGCA 300  
Dh 2238 TGCCTGACCGCTCTTCTTCTGCGATCTGCGCTGCGAGCGGCTGCGAGTACCTGCA 2297  
Qy 301 CATGACTTCTGCTGAGAGTGTCTGAGAGCGCTGAGAAACCGCGGATGCTCTCATCG 360  
Dh 2298 CATGACTTCTGCTGAGAGTGTCTGAGAGCGCTGAGAAACCGCGGATGCTCTCATCG 2357  
Qy 361 GCGCTGCTGCTGAGAGAGAGAGTGTCTGAGAGTGGGAGTGGGCGGCTGCGAAATGAGGCC 420  
Dh 2358 GCGCTGCTGCTGAGAGAGAGAGTGTCTGAGAGTGGGAGTGGGCGGCTGCGAAATGAGGCC 2417  
Qy 421 GCGCTGCGGAGAGAGAGCGGCTGCGAGAGTGAAGTGAAGCGGCGGCTGAGGAGCATG 480  
Dh 2418 GCGCTGCGGAGAGAGAGCGGCTGCGAGAGTGAAGTGAAGCGGCGGCTGAGGAGCATG 2477  
Qy 481 CCGAGGCTGAGAGCGGCGGAGTGTGATGCTTGGGCTGAGAGGCGGCTGTGTGCG 540  
Dh 2478 CCGAGGCTGAGAGCGGCGGAGTGTGATGCTTGGGCTGAGAGGCGGCTGTGTGCG 2537



Qy	541	GTGATATACGGCTATTTGCTCTCCGTGGTGGCCCTTATCTGAGCTGACAGATTTGCG	600
Dp	2538	GTGGATACGGCTATTTGCTCTCCGTGGTGGCCCTTATCTGAGCTTGGACGCTTGCG	2597
Qy	601	TCCGGAGATATGCTCCACGGCCCTGTGGTGGGATTCGTAGTGTTCGCCAGACACC	660
Dp	2598	TCCGGAGATATGCTCCACGGCCCTGTGGTGGGATTCGTAGTGTTCGCCAGACACC	2657
Qy	661	CTCGTGTGCTCTCGAAGATCCCGGCGCTGGCCACGGACGGTCCGTGCAMGGCATTTTCG	720
Dp	2658	CTCGTGTGCTCTCGAAGATCCCGGCGCTGGCCACGGACGGTCCGTGCAMGGCATTTTCG	2717
Qy	721	GCGAGGGCCGATGGGATTTCCGACGAGGGGAGCGCCGTCGAGTCTCAACGGCTC	780
Dp	2718	GCGAGGGCCGATGGGATTTCCGACGAGGGGAGCGCCGTCGAGTCTCAACGGCTC	2777
Qy	781	AGTGAAGCCCGCGCGGACCGCGATTCGATATTGGCGGTGATTCGAGATCCGCATCAAT	840
Dp	2778	AGTGAAGCCCGCGCGGAGCGCGATTCGATATTGGCGGTGATTCGAGATCCGCATCAAT	2837
Qy	841	CACGACGGTCCGACGACGCGTCTGAACGCTGCCAACCGGAGCTCCCAAGAAATGTGCTG	900
Dp	2838	CACGACGGAGCGACGAGCGGTCTGAACGCTGCCAACCGGAGCTCCCAAGAAATGTGCTG	2897
Qy	901	AAAGGGGCGCTGGGGAGACGAGGGCTGGCGCGCTTCCTCGAGGGATTATGACAGGACAC	960
Dp	2898	AAAGGGGCGCTGGGGAGAGCAGGGCTGGCGCGCTTCCTCGAGGGATTATGACAGGACAC	2957
Qy	961	GGCACGGGACGACGCGTGTGTGAACCCCATCCAAATCCAACTCGAATGCGGTATACGGC	1020
Dp	2958	GGCACGGGACGACGCGTGTGTGAACCCCATCCAAATCCAACTCGAATGCGGTATACGGC	3017
Qy	1021	CTCGGGCGAATGTTCGCCACGCGCTGCTGATCGGATGGTGAAGAACCAACTTGGCCAT	1080
Dp	3018	CTCGGGCGAAGCTTCGCCACGCGCTGCTGATCGGATGGTGAAGAACCAACTTGGCCAT	3077
Qy	1081	CCTGAGTATGCGTGGGGATCATCTGGGCTGCTGAAGAGCGTCTTCCCTTCACACGGG	1140
Dp	3078	CCTGAGTATGCGTGGGGATCATCTGGGCTGCTGAAGAGCGTCTTCCCTTCACACGGG	3137
Qy	1141	CAGATTCTCTGGCGACCTTCACGCGCAGGCGCTGAACCCCGGATCTCAATGGGGATCTT	1200
Dp	3138	CAGATTCTCTGGCGACCTTCACGCGCAGGCGCTGAACCCCGGATCTCAATGGGGATCTT	3197
Qy	1201	CGGCTGACCGTTCACGCGCGCCCGGACACCTGTGGCTCGACTTGAATACGCCCGCACGGGCG	1260
Dp	3198	CGGCTGACCGTTCACGCGCGCCCGGACACCTGTGGCTCGACTTGAATACGCCCGCACGGGCG	3257
Qy	1261	GGGGTGAAGCTGTTCCGCGATGAGCGGGGACCAACGGCGCACGTGGTGTGGAAGAGCGCGG	1320
Dp	3258	GGGGTGAAGCTGTTCCGCGATGAGCGGGGACCAACGGCGCACGTGGTGTGGAAGAGCGCGG	3317
Qy	1321	GCGGCGACGTGCACACCGCGCGCGCCGGAGCGACCGGACAGACTTGTGGTCTGTCCGCA	1380
Dp	3318	GCGGCGACGTGCACACCGCGCGCGCCGGAGCGGACCGGACAGACTTGTGGTCTGTCCGCA	3377
Qy	1381	AGGACCGGCTCAAGCCCTTGGATATGCAACGAGCGCGGCTGCGCCGACATCTTGGAACTTAC	1440
Dp	3378	AGGACCGGCGAGCCTTGGATATGCAACCGCGCGGCTGCGCCGACATCTTGGAACTTAC	3437
Qy	1441	CCTTCGCAAGTCTTGGGCGATATGTGGCGTTCAAGTCTTGGCGACGACCGCGCAGATGGAG	1500
Dp	3438	CCTTCGCAAGTCTTGGGCGATATGTGGCGTTCAAGTCTTGGCGACGACCGCGCAGATGGAG	3497
Qy	1501	CACCGGCTCGCGGTGGCGGCGACGTTCAGGGAGGGGCTGCGGCGAGCCCTTGAACGCTCGG	1560
Dp	3498	CACCGGCTCGCGGTGGCGGCGACGTTCAGGGAGGGGCTGCGGCGAGCCCTTGAACGCTCGG	3557
Qy	1561	GCGCAGGGACAGACGTCGCGCCCGGTCGCGGTGGCGAGTACCGCGAATTCCTCACGCGGCAAG	1620
Dp	3558	GCGCAGGGACAGACGTCGCGCCCGGTCGCGGTGGCGAGTACCGCGAATTCCTCACGCGGCAAG	3617

QY	1621	TTCCGCTTTCTCTCACCGGACAGAGGGGCGAGACGCTGGGCAATAGGGCCGGTGGCTGTAC	1680
Db	3618	CTCGCTTTCTCTTACCGGACAGAGGGGCGAGACGCTGGGCAATAGGGCCGGTGGCTGTAT	3677
QY	1691	GATGATATGTCGCGGTTCCGGAGGCGTTCGACCTGTGTGCGTGAAGCTGTTCAACAGAG	1740
Db	3678	GATGATATGTCGCGGTTCCGGAGGCGTTCGACCTGTGTGCGTGAAGCTGTTCAACAGAG	3737
QY	1741	CTTCGACCGGGCGCTTCGGCGAGGTGATATGTGGCGCGAACCGGCAAGCGTTCAGACCGCGGCTG	1800
Db	3738	CTTCGACCGGGCGCTTCGGAGGTATGTGTGGCGCGAACCGGCAAGCGTTCAGACCGCGGCTG	3797
QY	1801	CTTCGACGAGACGCTTCACCCAGCCGGCGCTGTTCACCTTCGATATATGGCTTCGCCGCG	1860
Db	3798	CTTCGACGAGACGCTTCACCCAGCCGGCGCTGTTCACCTTCGATATATGGCTTCGCCGCG	3857
QY	1861	CTGTGGCCGTCGTGTGGGCTGTAGACCGCGAGTTGTCGCCCGCATATGATCGGTAGCTG	1920
Db	3858	CTGTGTGGCCGTCGTGTGGGCTGTAGACCGCGAGTTGTCGCCCGCATATGATCGGTAGCTG	3917
QY	1921	GTCGCTGCTGTGCGGTGGGGGGGCGTATTCGCTTCGATGAGACGGGGGTCCTGTGTGCTGCG	1980
Db	3918	GTCGCTGCTGTGCGGTGGGGGGGCGTATTCGCTTCGATGAGACGGGGGTCCTGTGTGCTGCG	3977
QY	1991	CGCGGGCGCTGTATGCAAGCGCTTCGCGCGCGGGGCGATGTGTGATTCAGAGCGCGCG	2040
Db	3978	CGCGGGCGCTGTATGCAAGCGCTTCGCGCGGGGCGATGTGTGATTCAGAGCGCGCGCG	4037
QY	2041	GAGGCGCATATGTGGCTGTGCGGTGGGGCGCGGACGACGTCGTGGGTGTATTCGCGCGGTCG	2100
Db	4038	GAGGCGCATATGTGGCTGTGCGGTGGGGCGCGGACGACGTCGTGGGTGTATTCGCGCGGTCG	4097
QY	2101	AACGCTCCGAGCAAGTGTGTATCGCGAGCGCGCGGCGAACCCGTTCGATTCGATCGCGAG	2160
Db	4098	AACGCTCCGAGCAAGTGTGTATCGCGAGCGCGCGGCGAACCCGTTCGATTCGCGAGCG	4157
QY	2161	GCGATGCGCGCGCGCGGGGGCGCGAACCAAGGCGCTCCACGTCCTGGCATTCGCTTCACATTC	2220
Db	4158	GCGATGCGCGCGCGCGGGGGCGCGAACCAAGGCGCTCCACGTCCTGGCATTCGCTTCACATTC	4217
QY	2221	CGCGCTCATGCGCGCGGATCTGGAAGGCGGTTCGGGCGGTGTGGCGGATGCGTATCGCG	2280
Db	4218	CGCGCTCATGCGCGCGGATCTGGAAGGCGGTTCGGGCGGTGTGGCGGATGCGTATCGCG	4277
QY	2281	CGGCGTCGATCGTTCGTGTGACGAATTCGAGCGGGAAGCGTTCACAGACGAGGTGAC	2340
Db	4278	CGGCGTCGATCGTTCGTGTGACGAATTCGAGCGGGAAGCGTTCACAGACGAGGTGAC	4337
QY	2341	TGCGCGGGCTATTTGGGTGCGCGACGCGCGAGAGGTGGTGGCGCTTCGGGGATGAGGTGAAG	2400
Db	4338	TGCGCGGGCTATTTGGGTGCGCGACGCGCGAGAGGTGGTGGCGCTTCGGGGATGAGGTGAAG	4397
QY	2401	GCGCTGCAAGCGGCGCGGTGCGGCGACCTTCGTTCGAGGTTCGGTTCGGAATTCGACGCTCTC	2460
Db	4398	GCGCTGCAAGCGGCGCGGTGCGGCGACCTTCGTTCGAGGTTCGGTTCGGAATTCGACGCTCTC	4457
QY	2461	GCGCTGTGTGCTGCTGCTGATATGCGGACGCGCCCGCGCGCGCTGCTTCGATGCTGCGCGCTG	2520
Db	4458	GCGCTGTGTGCTGCTGCTGATATGCGGACGCGCCCGCGCGCGCTGCTTCGATGCTGCGCGCTG	4517
QY	2521	GGGCGTGAAGACCGGCGACGCTGTCTGAGAGCGCTTCGCGCGGCTTCGTGGGCGGTGCGTGTGC	2580
Db	4518	GGGCGTGAAGACCGGCGACGCTGTCTGAGAGCGCTTCGCGCGGCTTCGTGGGCGGTGCGTGTGC	4577
QY	2581	CTGTGTCTCTGTGGCCGGCTTCTTCCTCAGGGGGGCGGCGGGTGTGCTGCTGCCACGTAC	2640
Db	4578	CTGTGTCTCTGTGGCCGGCTTCTTCCTCAGGGGGGCGGCGGGTGTGCTGCTGCCACGTAC	4637
QY	2641	CCTTGTGACGCGAGACGCTATCTGGATTCGACACGAAAGCGGACGAGCGCGCGTGTGGCGAC	2700
Db	4638	CCTTGTGACGCGAGACGCTATCTGGATTCGACACGAAAGCGGACGAGCGCGCGCGTGTGGCGAC	4697
QY	2701	CGCCGTCGTCGCGAGCGGATCACGACGAGTTCGAGAGGGGGGCGCGGTTCGCGCGCGCG	2766

Db	4658	CGCCCTGCTCCGGGAGCGGGTTCACGACGAGTCCAGAAAGGGGGGCGCGGTCCGCGCGGC	4757
OY	2761	GACCGCGCAGCGCTCGGCTCGACCATCGCGCCGCCGAGCGGACCGCGGAGAAAGTC	2820
Db	4758	GACCGCGCAGCGCTCGGCTCGACCATCGCGCCGCCGAGCGGAGCGCGCGGAGAAAGTC	4817
OY	2821	GAGGCGCGCGCGCAGCCGTCGCTCCGGCTCGAGATCGATGAGCCAGCGCGTCTGATCAC	2880
Db	4818	GAGGCGCGCGCGCAGCCGTCGCTCCGGCTCGAGATCGATGAGCCAGCGCGTCTGATCGC	4877
OY	2881	CTCGGTCTCGGGATCGAGAGCGGCGCGCCCTGGTCTGAGGGCGAGGTGAGATGCGCGTC	2940
Db	4878	CTCGGTCTCGGGATCGAGAGCGGCGCGCCCTGGTCTGAGGGCGAGGTGAGATGCGCGTC	4937
OY	2941	GACGCGCGCGGGCTCAGCTTCATGATGTCCAGCTCGCGCTGAGGCAATGATGATGCCGACAC	3000
Db	4938	GACGCGCGCGGGCTCAGCTTCATGATGTCCAGGCTCGCGCTGAGGCAATGATGATGCCGACAC	4997
OY	3001	CTGCGCGGAAAGCCCAACCTTCGCTCGCTCGAGAGCGATGTGCCCGGCGCATGTC	3060
Db	4998	CTGCGCGGAAAGCCCAACCTTCGCTCGCTCGAGAGCGATGTGCCCGGCGCATGTC	5057
OY	3061	GCCGTGGGCGAGGGGCGTAAAGGCGCTGTGGTGGGCAACCGGTCATCGCCCTTCGGCG	3120
Db	5058	GCCGTGGGCGAGGGGCGTAAAGGCGCTGTGGTGGGCAACCGGTCATCGCCCTTCGGCG	5117
OY	3121	GGAAGGTTCGTACCCAGTCACGACGACGTCGCGCTGCGCTGAGTGTGCTCGGCTCAGAGG	3180
Db	5118	GGAAGGTTCGTACCCAGTCACGACGTCGCGCTGCGCTGAGTGTGCTCGGCTCAGAGG	5177
OY	3181	CTCTCGCGGATCGAGGCGGCGCCCATGCGCTCGCTGATCTTGACGCAATGATCGGCTC	3240
Db	5178	CTCTCGCGGATCGAGGCGGCGCCCATGCGCTCGCTGATCTTGACGCAATGATCGGCTC	5237
OY	3241	GACGAATAGCGCGCTTCAGCGCGGGGAGGGGGTGCATCCATGCGGCGCACCGGGGGG	3300
Db	5238	GACGAATAGCGCGCTTCAGCGCGGGGAGGGGGTGCATCCATGCGGCGCACCGGGGGG	5297
OY	3301	GTCGGTCTCGCGCGGATGCAATGAGGCGGACGACGTGGAGCGCGAGTTCATGCGACGAGC	3360
Db	5298	GTCGGTCTCGCGCGGATGCAATGAGGCGGACGACGTGGAGCGCGAGTTCATGCGACGAGC	5357
OY	3361	GGCACGCGCCGAGAAACGCGCCTTACTGTGAGTCTGCTGGGCGTGGGTATGTAGCGAATCC	3420
Db	5358	GGCACGCGCCGAGAAACGCGCCTTACTGTGAGTCTGCTGGGCGTGGGTATGTAGCGAATCC	5417
OY	3421	CGCTTCGACCGGTTCTGTGCGCGACGTGGCGCGCGCTGTGAGACGGGCGAGAGGTGAGCGTC	3480
Db	5418	CGCTTCGACCGGTTCTGTGCGCGACGTGGCGCGCGCTGTGAGACGGGCGAGAGGTGAGCGTC	5477
OY	3481	GTCGTCAACTGCGCTCTCGGCGGAGCTGATTCACAAGAGTTTCAATCTCTTCGATTCGAC	3540
Db	5478	GTCGTCAACTGCGCTCTCGGCGGAGCTGATTCACAAGAGTTTCAATCTCTTCGATTCGAC	5537
OY	3541	GCGCGGTTTGTGAGACTCGGCAAGCGGACTGTTAACGCGAATAACAGCTCGGCGTGG	3600
Db	5538	GCGCGGTTTGTGAGACTCGGCAAGCGGACTGTTAACGCGAATAACAGCTCGGCGTGG	5597
OY	3601	CCGTTTCTCGGCGCAATCTCTCTCTCTCTCGTGGTGGATCTTCGCGGGGATGATGCTGAGCGG	3660
Db	5598	CCGTTTCTCGGCGCAATCTCTCTCTCTCTCGTGGTGGATCTTCGCGGGGATGATGCTGAGCGG	5657
OY	3661	CCGCGCGCGGATCGGTCGCTCTTGAGAGAGTCTCGGCGCTGATCGCGGCGAGCGGTGTC	3720
Db	5658	CCGCGCGCGGATCGGTCGCTCTTGAGAGAGTCTCGGCGCTGATCGCGGCGAGCGGTGTC	5717
OY	3721	ACCCCTCCCCCATGCGACGCTCCCGATCGCCCGGTGTCCCGATGCGTTCCGAGCATG	3780
Db	5718	ACCCCTCCCCCATGCGACGCTCCCGATCGCTCGTGTCCCGATGCGTTCCGAGCATG	5777
OY	3781	GCGCAGGCGGCAATCTTGGAGTGTGATCTCACGCTGGGTGACCCGAGGTCCAATC	3840

[illegible]

Db 254 GCTCGTGGCGATCCACCTTGCTGCGCGAGCGCTGCGGAGCGATCTCGCGCT 313  
Qy 624 GGCTGTGGGATATCGCTGATGTGTGCGCGAGCAACCTGTGTGGCTCTCGAAGACCGG 683  
Db 314 GGCGGAGGCGGCAATGCTCTCTTTCGTCGAAGACGATGATATGCTGGGCGCATCCA 373  
Qy 684 GGCGGTGGCGAGGAGCGGTGCTGCAAGGCAATTTTCGGGAGGCGCGATGGGTTTGGAGC 743  
Db 374 GGCGGTGTGCGCGGATGGCACTGCGCGACATTCAGCGCTGCGCGAAGGGGTTTGTGTGG 433  
Qy 744 AGCGCAAGGATGCGCGCTGCTGATGCTCTCAAGCGGCTCAGTGGAGCCCGCGCGAGCGCGA 803  
Db 434 TGGGAGAGGCTTCGGATATGTGTGTCTCAAGGCTCTCCAGCGCCAGCGAGCGAGCGA 493  
Qy 804 TCGGATATTGGCGGTGATTCGAGATCCCGATCAATCAAGCGGTGCGAGCGCGGTCT 863  
Db 494 TCGGATCTGGGCTGTGATCCGGGGTTCGGCATGATGAGATGCGGTGCGAGCGGT 553  
Qy 864 GACCGTGGCGAGCGGAGCGCTCCAGAGAAATCGTGTGAGAAAGGGCGCTGGCGGAGCGAG 923  
Db 554 GATGCAACCCAAATGTGCTGCTCAGAGAGCGCTTTCGCGAGCGCGCTGCGAAGCGCTCG 613  
Qy 924 CTGCGCGCGCTTTCGTTGGGTTATGTGAGGCAACCGGCAAGCGGCAACCGCTTGTGTA 983  
Db 614 CGTCAACGCGGGGCGCATGCTTATGTGAGACCAAGGAGCGGAGCGCTCGCTGCGCGA 673  
Qy 984 CCCCATCGAAATCCAAAGCTTGTGATGCGGTATACGGCTTCGGGCGAGATGTGCGACGCG 1043  
Db 674 CCCGATCGAGGTGAGGCGCTGCTGCGCTGTGGGGGCGGCGCGGCGAGTGGAGCGCG 733  
Qy 1044 GCTGCGATCGGGGTGAGTGAAGACCAACTTGGCCATCTGTGATATGCTTGGGGATCC 1103  
Db 734 CTGCGGTGCGGCGGAGTGAAGACCAACTTGGCCATCTGTGAGGGGCGCTGCGAGGGGTGCG 793  
Qy 1104 TGGGCTGTGAAGGTGCTTGTCTTCCCTTCAAGCAAGGCGAGATTCTCTGCGACCTCCAGC 1163  
Db 794 GGGTTTGAATCAAGCGGGGCGCTGTGTGACCAAGGATGCTCCGCGAAACTTCCATT 853  
Qy 1164 GAGGCGCTGAACCCCGGATTCATGGGGTATCTTCGGCTGACCGTCAAGCGCGCGCG 1223  
Db 854 CCACACGCTCAATCCGCGGATCCGAGTCAGGGGAGACCGCGCTCGCGTGGCGACGAGCGC 913  
Qy 1224 GACACGTTGGCGGAGCTGGAATACCGCGCGAGCGGGGGGAGAGTGGTGGGATGAG 1283  
Db 914 GGTGCGTGGCGCGGCGGCGCGAGACCGCGCTTCGGGGGGTGAAGCGCTTGGCGCTCAG 973  
Qy 1284 CGGAGCAACGCGCAGTGTGCTGGAAGAGCGCGCGGCGAGCGATGTCACACCGCGCGC 1343  
Db 974 CGGAGCAACGTCATGTGTGTGTGAGAGGCGCGCGCGAGCGTCTCGACCGGCGAC 1033  
Qy 1344 GCGGAGCGAGCCGCGAGAGCTGTGTGTGTGCGGAGAGACCGGCTCAGCCCTGAGTGC 1403  
Db 1034 GCGCGGGGCGCTCAGCGGAGCTTTTGTGTGTGCGGAGAGACCGCGCGCGCTGAGACCGC 1093  
Qy 1404 ACAGGCGGCGGCGGCTGCGCGAGCATCTGAGAGCTTACCTTGGCACTGTCTGGGCGATG 1463  
Db 1094 ACAGGCGGCGGCGGCTTCTCAGCGGCAATCGCGCGTACCGGAGCAGAGGCTTGGAGACGT 1153  
Qy 1464 GCGCTTCAATCTGCGAGCAGCGCGAGCGAGTCGAGTGAAGCAACCGCTTCGCGGTGGCGAGC 1523  
Db 1154 GCGCTTCAAGCTGTGATGAGCGGTAAGCGCGATGAGAGCAACCGGCTTGGCGGTGGCGAGC 1213  
Qy 1524 GTGAGAGGAGGAGGCTGCGGAGCGCTGAGCGCTGCGGCGAGAGGAGCAGACGTCGCGCG 1583  
Db 1214 CTGCGCGAGGCGCTGCGAGCGCTGAGGTTGGGCGGCGAGGAGCAGACCGCGCGAGG 1273  
Qy 1584 TGGCGGTGCGCAATATCGCGGATTCCTCAAGCGGCAAGCTGCGCTTCTTCAACGAGAA 1643  
Db 1274 CGCGGCGCGCGAGGCGCGCTTCTCGCGCGGCAAGCTCGCTTCTTGTGCGGGGGA 1333  
Qy 1644 GGGGCGCAGACGCTGGGAGTGGGCGGTGAGCTGTATGATGTATGCTGCGCTTCCGCGA 1703  
Db 1334 GGGCGCGAGTGGCGGGGATGGGCGGTGTGGGAGGCGGTGCGCGGCTTCCGCGA 1393

Qy 1704 GGCGTTGCACTGTGCGTGAAGGCTGTTCAACAGAGGCTCGACCGGCGCGCTCCGAGAGT 1763  
Db 1394 GACCTTCAACCGGTGCGTCAAGCTCTTTCAGCGGAGGAGCTTCATACGCGCTGTGGAGGT 1453  
Qy 1764 GATGTGGGCGAGCCGCGCAGGCTGAGCGCGCGCTGCTGACGAGCAGGCTTCAACCA 1823  
Db 1454 GATGTGGGCGAGCCGCGCAGGAGCAGAGGTGCTGCTGCTGAGACAGAGGCTTCAACCA 1513  
Qy 1824 GCGGCGGCTGTTCACCTTCAATATGCGCTGCGCGCGCTGTGAGCGGTGTGAGA 1883  
Db 1514 GCGGCGGCTCTTGTGCGTGAAGTACGCTGCGCGCGCTTCTTCGCGGTGTGGGGGTGGA 1573  
Qy 1884 GCGGAGTTGGTTCGCGGCGCATAGATGGGTGAGTGTGCTGCTGCGCGGTGCGGCGGT 1943  
Db 1574 GCGGAGGCTGTGTGCGTCAATAGCTGTGAGCTGTGAGCTGTGCGCTGTGCGGCGGT 1633  
Qy 1944 GTTCTGCGTTGAGAGCGGAGTGTCTGTGTGAGTGTGCGGCGCGCTGATGCAAGCGCT 2003  
Db 1634 GTTCTGCGTTGAGAGCGGAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1693  
Qy 2004 GCGGCGCGGCGGCGAGTGTGTGATGAGCGCGCGGAGCGCGAGTGTGTGTGTGTGTGT 2063  
Db 1694 GCGGCGCGGCGGCGAGTGTGTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1753  
Qy 2064 GCGGCGCGAGCGAGT 2123  
Db 1754 GCGGCGCGAGCGAGT 1813  
Qy 2124 GCGGCGCGCGGAGCAACCGGTGATGCGATGCGGCGGCGAGTGTGTGTGTGTGTGTGTGT 2183  
Db 1814 GCGGCGCGCGAGAAATTCGTGTGAGAGATGCGGCGGCGGTTCGCGGCGGCGGCGCG 1873  
Qy 2184 AACCAAGCGCTCAAGT 2243  
Db 1874 AACCAAGCGCTCAAGT 1933  
Qy 2244 GCGGCTGCGGCGGT 2303  
Db 1934 GCGGCTTCCGCGGAGT 1993  
Qy 2304 CAATGTAGCGGAGGAGCTTCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2363  
Db 1994 CAATGTAGCGGAGGAGCTTCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2053  
Qy 2364 GCGCGAGAGGT 2423  
Db 2054 GCGCGAGAGGCGGT 2113  
Qy 2424 CACCTTGT 2483  
Db 2114 CCTTGT 2173  
Qy 2484 GAGCGCGCGCGCGGCTGT 2543  
Db 2174 GAGTGTCAAGCGGAGT 2233  
Qy 2544 GCTGAGGCGCTGCGCGGCTGT 2603  
Db 2234 GCTAGAGGCGGCTGT 2293  
Qy 2604 CCGCTCAAGGGGCGGCGGAGT 2663  
Db 2294 CCGCTTGGGCGGAGCGGCGGAGT 2353  
Qy 2664 GATGCAACAGAAAGCGAGCAGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2723  
Db 2354 GATGCAAGCGCGGT 2413  
Qy 2724 CGAGAGGTGAGAGG 2740  
Db 2414 CCGCTTCTGGGTGAAG 2430

RESULT 11  
AX024276 10910 bp DNA Linear PAT 24-NOV-2000  
LOCUS Sequence 81 from Patent DE19846493.  
DEFINITION AX024276 AX024383  
ACCESSION AX024276.1 GI:10184550  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Polyangium cellulosum  
Polyangium cellulosum  
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
Sorangineae; Polyangiaceae; Polyangium.  
REFERENCE  
1  
AUTHORS Beyer, S. and Mueller, R.J.  
JOURNALS Patent: DE 19846493-A 81 13-APR-2000;  
BIOTECHNOLOG FORSCHUNG GMBH (DE)  
COMMENT On Oct 15, 2002 this sequence version replaced gi:10184587.  
FEATURES  
source  
1. 10910  
/organism="Polyangium cellulosum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:56"  
BASE COUNT 1852 a 3124 c 3613 g 2321 t  
ORIGIN  
Query Match 21.8%; Score 928.8; DB 6; Length 10910;  
Best Local Similarity 60.5%; Pred. No. 3,36-92;  
Matches 1606; Conservative 0; Mismatches 1027; Indels 21; Gaps 4;  
QY 25 GCACCCGAAGATCCGATTCGATCGTGGAGCGATTGCCCTGCGCGGTGCGGTATC 84  
DB 2244 GCGCGGATGACGATCGCATCGCTGCTGCCCTGCGGTTCGCCAGGTGGAGATGAG 2303  
QY 85 GATCTAGCGGGATCTGAGAGCTCTCGAGGGCTCGGCGACACCTGCGGGCAGTCCC 144  
DB 2304 GCGCTGAGACATCTGCGGCGATCTGCGCGAGGAGCATGTGTACGACCGAGGTGCA 2363  
QY 145 GCCGACGCTGGAGTGAAGAGCGGTGTTGATCCCGACCCGAGTGCCTCGGGAGAGC 204  
DB 2264 GCGACGCGTGGCGCGGGCGGAGCTGTAGACCCCGATCCGAGGTTCGCGGCGGAGC 2423  
QY 205 CCGCTTACGCGCGCATCTTCTCTAGAGCAGTACCTGCTTGAAGCGCTCTTCTTGGC 264  
DB 2424 TATCTGCGCAAGGCTGCTCTCTCGGATGTGCGAGCTTGAATGCGGCGCTTCTGCGC 2483  
QY 265 ATCTGCGCTGCGGAGCGCTGCGAGTGAACCTCGACATCGACTTCTGCTGAGGTGCG 324  
DB 2484 ATTTCCTCTGAGGCGATAGCTTGACCGGACCGAGCGGCTGTGTGAGGTGAGC 2543  
QY 325 TGGAGCGCTGAGAAAGCGCGCATCGCTCATCGGCGCTGTCGCTGAGTACGAAACGGA 384  
DB 2544 TGGAGGCGATGAGAGCGCGCTGCGAGAACCCGATGCGCTGCGGAGAGCGCACCGGC 2603  
QY 385 GTGTTCATCGGAGTGGCGCGCTCGGAATAGAGCGCGCTGCGGCAAGCGAGCGCTCC 444  
DB 2604 GTGTTCGTTGCGATGATGAGGAGCG--AGCAGCGCGAGCGGATGAGAGGCGCTGAGAC 2660  
QY 445 GCAGAGATCGACGCTCATGTGGGGCTGGGAGCATGCGCCAGCTGCGAGCGGGCGGATC 504  
DB 2661 GACCGCGCTGCTGTAAGGACCAACCGCAACTGTCTACAGCTGCGCGCTGAGACGCGCTG 2720  
QY 505 TCGATATCCCTCGGAGCTGCGAGGCGCGATGTGTCGCGGAGTACGAGCTATTGCTCCG 564  
DB 2721 TCGTTCCTCTGCGGTCTGCGAGCGCGCCGACGATGACGCTGAGACGCGCTGCTGCTG 2780  
QY 565 CTGGTGGCGCTTCACTGCGCTGCAAGCTTGTGCGCTCGGGAGATGCTCCACGCGCTCG 624  
DB 2781 CTGGTGGCGCTTCACTGCGCTGCAAGCTTGTGCGCTCGGGAGATGCTCCACGCGCTCG 2840  
QY 625 GCTGCTGGGAGTATGCTGATGCTGCGCGGAGCAGCTGCTGCTGCTGCTGCTGCTGCTG 684  
DB 2841 GCGCGCGGCTCAGCGCTGCTTGTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2900

QY 685 GCGCTGCGCAGGACGGTGTGCTGCAAGCATTTTTCGCGGAGCGGATGAGTTCGAGCA 744  
DB 2901 TTGCTTTTCCCAAGATGAGCGGTGCAAGACGTTCTTCGCGCGCTGAGACGCTTTCGCGG 2960  
QY 745 GCGCAAGGTGCGCGCTGCTGCTGCTTCAACGCTGCTGAGTGGAGCGCGCGGAGGAGT 804  
DB 2961 GCGGAGGCTGCGCGCTGCTGCTGCTTCAACGCTGCTGAGTGGAGCGCGCGGAGGAGT 3020  
QY 805 CGGATATGCGCGGTATTCAGAGATCCGATCAATCAAGCAGGTGCGAGCGGTCTG 864  
DB 3021 CCATCTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3080  
QY 865 ACCGTGCGAAGCGGAGCTTCCAGAAATGCTGTAAGCGGCGCTGCGGAGCGAGCG 924  
DB 3081 ACGGTGCGCAGCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3140  
QY 925 TCGCGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 984  
DB 3141 GTGCGCGCGCGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3200  
QY 985 CCATCGAATTCGAAGCTTGAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044  
DB 3201 CCGATCGAGGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3260  
QY 1045 CTGCTATCGGCTGCTGTAAGACCAACCTTGGCCATCTGATGCTGCTGCTGCTGCTGCTG 1104  
DB 3261 CTGCTATCGGCTGCTGTAAGACCAACCTTGGCCATCTGATGCTGCTGCTGCTGCTGCTG 3320  
QY 1105 GGGCTGCTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1164  
DB 3321 GGGCTGCTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3380  
QY 1165 CAGGCGCTGAACCCCGGATCTCATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1224  
DB 3381 GAGGCTTCAACCCCGGATCTCATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3440  
QY 1225 ACAACGTGCGGAGCTGGAATTAAGCGCGAGCGGCGGAGTGAAGCTGCTGCTGCTGCTG 1284  
DB 3441 GTCCCTGCG 3500  
QY 1285 GGGACCAAGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1344  
DB 3501 GGGACCAAGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3560  
QY 1345 CCGGAGCGAGCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1404  
DB 3561 CCGGAGCGAGCGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3620  
QY 1405 CAGGCGCGCGGCTGCGGACCATCTGAGACCTTTCGAGTGTCTGCGGAGTGTG 1464  
DB 3621 CAGGCGCGCGGCTGCGGACCATCTGAGACCTTTCGAGTGTCTGCGGAGTGTG 3680  
QY 1465 GCGTTCAGTGTGCGAGCAGCGCGAGCGCGATGAGACCGGCTGCGGTGCGGCGAGC 1524  
DB 3681 GCGTTCAGTGTGCGAGCAGCGCGAGCGCGATGAGACCGGCTGCGGTGCGGCGAGC 3740  
QY 1525 TCGAGGAGGAGGCTGCGGAGCGCTGAGAGCTGCGGCGGAGAGGAGCAGAGTGCCTG 1584  
DB 3741 TCGGCGAGGCGCTGCGAGGAGGCGCTTTCGCGCGAGCGGAGGAGCAGAGTGCCTG 3800  
QY 1585 GCGGTGCGAGTATCGCGCATCTTCTCAGCGG--GCAAGCTGCGCTTCTTCAACGGA 1641  
DB 3801 GCGGTGCGAGTATCGCGCATCTTCTCAGCGG--GCAAGCTGCGCTTCTTCAACGGA 3900  
QY 1642 CAGGCGCGCAGACGCTGCGCATGCGGCTGAGCTGAGCTGTAAGATGATGCTGCGCTGCG 1701  
DB 3901 CAGGCGCGCAGACGCTGCGCATGCGGCTGAGCTGAGCTGTAAGATGATGCTGCGCTGCG 3960  
QY 1702 GAGGCGCTTGAACCTGCGAGGAGGCTTCAACAGAGAGCTGCAACGCGCGCTGCGGAG 1761  
DB 3921 GCGGCGCTTGAACCTGCGAGGAGGCTTCAACAGAGAGCTGCAACGCGCGCTGCGGAG 3980  
QY 1762 GTGATGTGCGCGAAGCGCGCAGCGCTGCGAGCGCGCGCTGCTGCGAGACGCTTCAAC 1821



gene  
CDS

```

LASKVLAIVLHDTTPRASPAELERLEGMLSAIVADEALRDLTAARLAFLDRAVR
TERDDAFAFEKSGASADELIRLIDQAGRIDVRY"
7542..22595
/gene="sp1H"
7542..22595
/gene="sp1H"
/function="biosynthesis of spirangiene"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="polyketide synthase"
/protein_id="Cad3449.1"
/db_xref="GI:27528335"
/translation="MSTVTNDLTTEYLRLTQELHRSRTLRATEERHEPIAVIGLG
LRLPGHIDRLTMTFLEGRDALPIPSHRWNADATDPDPAVGSXYNDAMLR
VDLFDAPFGISPREAKYDPOHRLLETSMQLADAGIVPALSDDKQVFGVCGS
DYALFQSDRDASEAEMKMSIFRAKRLATLGLQGPALSDTPACSSVYALHACQ
SLRQECDLAVAGVQWMSPEVVLISRTALSDGSKTFSANADGREGCVVL
AVERLRARAKGRPIIAVIRGSANHDGSSGIVPNGAQKYLRAALDARLVPA
VDVECHGTGIDPIEVNLAAYVGGREGKDRPLFGALKTNIGHLEFASGLA
KMSVSMRATPLATLHTSPLNPLVDMDLPIRVVDARPMTRDDGAPRRAGVAGL
SGTAAHIVEEAPAEPTTPDAPALPAVVLISGKDEALRQAARLHAGLRD
ARLYDASLATTTHTRRAAVVADBDLGLDALRGEAGGSVVASITAGV
VVFPGSSQWGNARALASVGFRODIA CERLAIPHAMSJGAVLRGDDATLL
GRVDVQVPIFAVWVALAALMRSIGVTPDAVAGSOGELAAVYAGALSLEDAVVA
LRARLTKLAGRGAAMAVELGADTEARLAFGDAIAIAINSRATLVAGDTAIDA
LVRLEAQAIFARKVAVYASHAVEIERELADAGIEPRAGAVPLVAGVAGL
DGSRLDAAMFRNLRSTYRFEADTRALHDGHRFPVPPVLAALDELTSIGD
AAVYSIRDEGDELERELISAGELHARGHAIIDKAPFAPLGRVTLPTTAFORERN
LDAGADDDDEDAFGASAEETAFWRALERGVAAISDLAHVADGSRSLASLAMPAL
AMRSREOSTVGAMRYRVEWRPVSAASRGVAGTWLVAPAGVASDLSALMALTE
RGADVPLEAEADITAHDAASRLDAIAGRPARGVLTPLDTPADLGRPAAGL
ARSLAIOALAEILAEPIWITRGAVAGISREDSIPVOMAMGLRVSLEQBEA
GGVLDLPRAHADGALDPAITELRLVLAALGSRGDEILAPGGLPARRLARLDA
QREPREVTLVGTGALGANVAMLRHGAHIVLASRGAAPGABALAEITELG
ARVTAACDVADRRAIBELFAPAEPELVIAFHTAGVTDGVLSLPERLETVLG
KADARTLHELTKTRDLAAFLVSSAGVFNCAQASVAAANALDLAEERLALG
ATPASYAMAGAGMLTEPDVQGLRGGAPEIAVAAALDRDBETHVTIASIDW
ARFASVYASARORLVADLPARALRQPSSTEHDTALVATLRALSGEARLHV
SLVLAOTRAVIGHDRSLDRBGRFPDLGLDLMAVEIRLOKATGVKIPATTIPR
PSPRKAGFLDLPLGGAAPSTARPAPARADASDEITAVIGLIPGQVDVDA
FRLEGERDAVAPIPASRDVGAVPDPEKGTGYVGHAAMLDRVLPDGFPGISIP
REARHVPORHLLETAMQLEAATIPASLRDSTGVFGASDVAVLOSSEDELA
AYAMGPASAPAGLAEFTLGLQGPALSDTPACSSVYALHACQSLRQECDLAA
GVOMASPEVVLISRTALAPDGRSKTESANADGREGCVVLALERLDARANH
PVAVVGTAVNHGASSGITAPNGSQOKTLPALDARLAPDVVVECHGTGSL
GDPIEVALAVIGEGRAPERLRLGAVKTNIGHLEFASGLAVKIVASIRHGTPLA
TIHTNPNRPHIDMALPIRVVDAEPWERSDGPBRAGVAFGLSGTNAVIVEEAP
EPDAPKARGLPLSGALPEVLSAKSDAALAAQAAALDRDPDALVDVAAALAT
TRSQDFRRAIADHDGALVLALEGAAGVAPGTFVAKGADLAFLETFQGGORAA
MGRGLYAPFVRDALDVAVSHLRELDRLPDLVLPAPKGSSELASLIDREFNOPALF
ALRYALRIVEANKVTPDVLIGHSGVGEILAAHVAGVLSLADACTVVAARALMALP
RDGMVTVHATEAVLALLEGDRAEILAINASSTVIAVDVAIVARVAFKAR
KATLRVSHAESHMDPMLDAFRVAGLTFHPRIPIVSNVGTALABEIRAPDRW
VRHRSVVRPADGINTLEADGVSSFEIPEGPVLSALGESALAAAGVAPFALRQ
RADVDLTAALSLHVRGHRVDMAFPEFPDPTVAFQDRFLDPAALRDB
EAAAPABEDAEMRADVAGVAGLGAITLGAEDHLSARELRLTALSAMRABERSL
VDARVYVVKRPLPTAARDVAGTMLVTLTLAGADHLSARELRLTALSAMRABERSL
PAEDRVRLAVVAGLADAGEELGIVSFALDTELTATLTHALPAGLALTRLVQALG
DMGIEAPLMTVRGASVSGSRDLASAQSMGTGIRGVGLGHEBERNGVADVGGT
IDARLRLVTLTASRDEQDLARPTGIFRRLVRAPLGEHAPARVTVAGTSVLG
GTGALGHAVAFLAGRGEHLVLSRREDDAGALAELELGAARTVLAACVABR
ASVALLARLADAGDPVRAVYHAGVAGVAAALSTDDGAGVAAVAAVRGALHILL
GDRLEDAFVLPAAGAGVWGGCOGAVAAAGNALDALVEPRAAGLTKTSTAKWAGG
GMLADHADAHEVELRKGLVPMAPALAIALAGLDHGETAVTVASDARLRAFA
SARPRLLDLPEAQRAL EASAPAGAGSADALVEALRPISEABARLVELVA
LATAVRLALPADALPADATTARSSSHVAIDEPIATVGLGLPGLGVREDBSGPFR
LEQDQGVGPIPKDRWDADATVDDPEAKGSYVBAALDRVLPALPAGFISPREA
KQDORHRLLETTMOALEATVYPAISKOTSGVFCIGISDVLMOAGSEETDAYA
AMGTLSAPAGRLAFTTGLQGPALSDTPACSSVYALHACQSLRQECDLAAGVQ
VMTLAEPRVLLSRTFALLADGRSKTFSAAHAGYRGEGVVALLERLDARARV
AVVGTAVNHGASSGITAPNGSQOKTLPALDARLAPDVVVECHGTGALGP

```

```

1EVOALAAVYEGEBAERPLLGLAVKSNIGHLEASGLAVAKIVTALHEGALPATIH
SRPNPHEVDMDALHVRVVDAAELMARDEBTPRAGVAGCLSGTNAHAIIEEPRAR
RGAATADPEPPAVPVVVADESERLRAHQBRLRAHYRGDGLVDVAAASLATSRRH
EHRATVARDGLLEDAITAGLBSRSGAGVITGKGLALLFTGGGGRGAAMGRGL
YGAFSVFRDAPDAACSLDBREHLRPLDVVPAPRGSLAALLDRTTYTQTSPLAEVA
LFLRVEAMGVTPPHVLGHSIGELVAAHVAGVLSDECTVLAARLHMLPRDAAV
VAHVASHEVLLDLDAAGLDGRASIAVNAVPSSTVYAGREAVLAARAEALGRKT
SRLRSHAESHMDPMLDAFRVAGLTFHPRIPIVSNVGTARASNDLCSADYVW
RHARGTVRFADGLEALADGVSTLEIGHVAVLGLALHETELSEAVVFLPLRQGRD
VDITLTAALGALHARGHVMDAPFAPAPNARVALLPYAFORERFLLMPKXSRRAA
ADSNGRYPLIAGTRLDLPDGSVHTLDIGPQAPLADHSTYGRIVYGAFRLAVILAV
AESHWPSPLERDVEFRVRLTEESPEVAVRNHLIPAGDGGFATVATVQEGAM
TTHATATGAAPSAALRHPRAHEPRAGBPASAVARLAAFPFAFVWGPVKRMVR
QTHAGCDRTLVGYLAADGVPTDDAPLPGCLVDNAPVAGSAVGLRPAHNDSTSNV
PLPESIRLTVWYGAETFCWADFMREBIEPDARVIGDITLVMBEGASGSAERGT
PIATIBGLARBPADKELBELPARNLYALTWPQAPASPPRRMALLGDSIGIAT
GIADADYPDVLDALRDLGRAPCEDTYAVCLATRHGSDTDETELVSVTHAARALA
LUNAMLSDEBLASTVLTLQRAVATAGEDVLDLASAALGLVRSQAEERPAIIL
VDLDGSDASRRALPYAVAGLADAEPLQALRDLGRRLRLRVAAPTAANAAITS
ATPRALPESTVLTGCTGTLGALVARRLVRHGVKILVLSRGGPAPGATDAAEL
QARGASVLAADAAEPALARVLLAIPHDRPLVAVVAGTLDGVLSSWTPARLSA
VRAKVDAVNILOBORTSPRLRAVFLSSLSGVTSQAQSVAAANALDALARBRA
RGUPVALDMGYNAQTSGLTKHEQDLRIRARGGLRPLASDEALPDLALADARC
LVPAFADALRKSGLDSALRGLVAGPPARAGSAGASALKURLASLATERH
RMLDPTVADAVATVYVVASPLDIEPGRLOELGDSLMAIELRRLAAATGLRLSATL
LFPHTEBALAKITISGILHDEAPPESTSAAEIDRLGALASVANAERLIDELTRMK
ALISKMAAEPSDDGLARLDTPANDDELFRMIDHVAEAAAT"
22592..33133
/gene="sp11"
22592..33133
/function="biosynthesis of spirangiene"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="Cad3450.1"
/db_xref="GI:27528336"
/translation="MSAPPGTEALRDYLLKQTLDLORSOTRLHAESKAYEPIAIV
MSCRFPGASPELMOQLGGRVAVGSPASRMSVSEPRVDPADVTGKPALEGF
LQAGELPDAPFCVGSREALEMPQOQLLETTMEALAEATIDAEISQSGTVFVY
MNNYASHHJLGGDDAGSYSPASSVASVAGRIAYTTFHGPIYTVDTAGSSVVAHA
LACQALSGSCALALAGVYVMAITREFDALSPFGSDGRCKSSSADANGSWEBAG
MVLIERLSDQNRGHPILAVIRGSVAODKSOQLTPNGAQCRVIRLOLAARLLET
HDVLAESHGTGRLGPIEIANALATVYGGAPERPLMGLSISNTGHTQAAAGVG
IKTIVLARNGLIPRTLHAHAPSPHVMAGGLRLADAVPTWNGVRAVASFGL
SGTAAHIVEEAPAEPTTPDAPALPAVVLISGKDEALRQAARLHAGLRD
GASLVDAASLATTTHTRRAAVVADBDLGLDALRGEAGGSVVASVARRAP

```

Query Match 20.1%; Score 859; DB 1; Length 4800;  
 Best Local Similarity 57.6%; Pred. No. 7,7e-85;  
 Matches 1735; Conservative 0; Mismatches 1220; Indels 48; Gaps 9;

Qy	14	CCATGAGCGCGGCAAGCGAAGATCGATTGCGATGTCGAGCGAGTTCGCTGCCG	73
Db	12286	CCCCCGCGGACGCGAAGCGACGAGCCATCGGATGTCGCGGCTCGCCTCCAG	12345
Qy	74	GTGCGCGATCGATCGAGCGGGCTTCGAGACGCTCTCGAGGGGCTCGCGGACACCGCG	133
Db	12346	GCGGGGTGACAGAGTGCAGCGCGCTCTTCGCGTGCAGGGGAGGCGACGCGGCTG	12405
Qy	134	GCGCAGTCCCGCGCGAAGCGTGGGATCAGACGGCTGTTGATCCGACCCCGATGCC	193
Db	12406	CTCCGATCCCGCGGACGCGCGTGGGACGTCGCGCGCGGTTGATCCGGAACCCGAGGA	12465
Qy	194	CGGGAAGACGCCCGTTACGCGCGCATCTTCTCTGAGCAGCTAGCTGCTTCAACGCG	253
Db	12466	AGGGGAAGAGTACGTCGCGCACCGCGCATGCTCGACCGGTCGACACTCTTCAACGCG	12525
Qy	254	CTTCTTCGCGATCGCTCGCGAAGCGGTCGCGGATGGAACCTCGACATCGACTTTCG	313
Db	12526	GCTTCTTCGCGATCGCCCGCGAGGCGAGGACGTCGACCCCGACGACCGCTCTTGC	12585
Qy	314	TGAGGTGTGCTGGAGGCGCTGAGAGAACCGCGATCGCTCATCGGCGCTCTCGGTA	373



Db 12586 TCGAGACCGCGGTGGCAAGCGCTTCGAGCGGCGGACGATGTCCTCCGCTGCTCCGCACT 12645  
Qy 374 CGGAAACGGAGATGTTTCATCGGAGATCGCCCGTCCGATATAGAGCGCGCTCGCGCAAG 433  
Db 12646 CGACACCGCGGCTTTCGTCCGCGCGGCGCGAGCGACTACGCG-----TGCTCC 12696  
Qy 434 CGACGCGCTCCGCAAGATTCAGCGCTCATGCGCGGCTGGGAGACGATGCCACGCTCGAG 493  
Db 12697 AGAGACGCGCTCAGGACCGCGAGGCTTACGCGCATGAGGACCGCGCTTGTTGCGCG 12756  
Qy 494 CGGCGCGAATCTGATGCTTCGCGGCTTCGAGGCGCTGCTGCGGCTGAGTACGCGCT 553  
Db 12757 CAGGTCGCGTGGCGCTTACGCTCCGCTCCAGGCGCGCGCTCTCGATTCAGACCGCGT 12816  
Qy 554 ATTGCTCTCGTGGTGGCGCTTTCATCTGAGCTTCGACGCTTCCGCTCCGCGAATGCT 613  
Db 12817 GCTCGCTCGCTCGCTCGCGCTTCCACCTCGCTGCGCATGCTGCGCGCAAGGCGAGTGC 12876  
Qy 614 CCACGCGCTTGGCTGGTGGGATATGCTGATGTTGCTCGCGAGCACCTCTGCTGCTCT 673  
Db 12877 ACCTGCGCTCGCGCGCGCGCTTCAGGATGAGCTTCGCGCGAGGCTTCTGATGATCT 12936  
Qy 674 CGAAGACCGCGCGCTGCGCGAGGACGCTGCGTGCAGAGGCAATTTTCGCGAGCGCGAG 733  
Db 12937 CGCGACGCGCGCTTCCGCTCCGAGACGCGCGATTCAGAGCGCTTCTCGCGAGCGCGAG 12996  
Qy 734 GGTTCGAGCAGAGCGAGAGGCTCGCGCTGCTGAGCTTCAGAGCGCTCAGTGGAGCGCGG 793  
Db 12997 GGTTCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13056  
Qy 794 CGGAGCGCGATGCGATATTTGGGCTGATTCGAGAGATCCGCGATTCATCGACGAGTGC 853  
Db 13057 CCAGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13116  
Qy 854 GCAGCGCTTCGATCGCTGCGAGAGGCTCCAGAGAAATCGTGCAGAAACGCGCGCTG 913  
Db 13117 CGAGCGCGATCACCGCGCGCGAGCGCTCTCGCAGAGAGAGTCTCCGCGCGCGCTCG 13176  
Qy 914 CGGAGCGAGAGCTCGCGCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 973  
Db 13177 ACGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13236  
Qy 974 CGCTTGTGATACCCATTCGAAATTCGAAATTCGAAATTCGAAATTCGAAATTCGAAAT 1033  
Db 13237 CCTCGCGGATCCATTCGAGGCTCAAGGCTTCGCGCGCTTCATCGCGAGAGGCGCGCC 13296  
Qy 1034 TCGCACGCGCGCTGCTGATCGGCTGCTGAGAGACCACTTGGGCTCTGAGTATGCT 1093  
Db 13297 CCGAGCGCGCTTCCGCTCCGCGCGGCTGAGAGACCACTTGGGCTCTGAGTATGCT 13356  
Qy 1094 CGGAGATCATCTGGCTGCTGAGAGTCTGCTTCTTCAGACGCGGAGATTCCTGCGC 1153  
Db 13357 CGGCGCTCGCGCGCTGCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13416  
Qy 1154 ACCTCAGCGAGCGCTGAGACCCCGGATTCATGCGGCTGCTGCTGCTGCTGCTGCTGCT 1213  
Db 13417 CGATCCACCAACCGCGCAACCGCATGACTGAGTGGAGCGCTCCCGCTCCGCTGCTG 13476  
Qy 1214 CGCGCGCGCGACACCGTGGCG-----GACTGGAATACGCGCGAGCGGCGGAGTGA 1267  
Db 13477 TCGAGCGCGCGCGAGCGCTGAGAGCGGAGCGCGAGCTCTCGAGCGCGCGGCGGCT 13536  
Qy 1268 GCTGCTTGGCATGAGCGGAGCAACGCGCACGCTGCTGAGAGAGCGCGG-----CG 1323  
Db 13537 CGCGCTTGGGCTTCTCGCGAGAGAGCGCATGCTGCTGAGAGAGCGCGCGAGCGCG 13596  
Qy 1324 GCGAGCTGACACCGCGCGCGCGCGAGCGACCGGAGAGCTG-----TGTGCTGT 1375  
Db 13597 AGCGCCGAGAGCGCGCGCGCGCGCTGCTGCTGCGCGCGCGCTCCCTTCTGCTGCT 13656  
Qy 1376 CGGCAAGAGCGCGCTGAGCTGCTGATGACAGCGCGCGCGCTGCGCGAGCTGAGAG 1435

Db 13657 CGGCGAGAGCGAGCGCGCGCTTCGCGCGAGAGCGCGCGCTCCGCGATGCTCGGCA 13716  
Qy 1436 CCTACCTTTCGAGTGTTCGAGCGATGTCGCTTTCATGCTGCGCACGACGCGAGCGGA 1495  
Db 13717 GAGACACCGAGCGCGCGCTTCGCTGAGCTGTCGCGCGCGCTCGCGACGAGCGCTGCAAT 13776  
Qy 1496 TGGAGCACCGCGCTGCGTGGCGGAGCGTGCAGAGGAGAGGAGCTGCGGAGCGCTGAGAG 1555  
Db 13777 TTGACACCGCGCGCGAGTTCGCGCGAGCATTCAGGAGCGCTGCTGCGCGCTGAGAG 13836  
Qy 1556 CTGCGCGCAGGAGCAGACGTCGCGCGTGCCTGCGAGTATGCGCATTCCTACCGG 1615  
Db 13837 GCGTCGCGCGCGCGCTGCGCGCGCGAGCGTT-----GCTGCGCAAGAGCGCGCG 13890  
Qy 1616 GCAAGCTGCGCTTCTTTCACCGGACGAGGAGCGAGACGCTGAGGATGAGGCGCTGAGC 1675  
Db 13891 ACAGCTGCGCTTCTGTTACGAGCGAGGCGCGACGCGCTGCGATGAGGCGCGCGC 13950  
Qy 1676 TGTAGATATGATCGCGCTTCGCGAGGCGTTGACCTGAGCTGAGGCTGTTCAAC 1735  
Db 13951 TGTAGACGCGCTTCCGCTTTCGCGAGCGCTGAGATGCGTCCGAGCACTTCATC 14010  
Qy 1736 AGAGCTGACCGCGCGCTTCGCGAGGATGATGAGCGGAGACCGGCGAGCTGACGCG 1795  
Db 14011 GCGAGCTGATGCGCGCTTCGCGAGCGCTTCCTTCGCGCGAGGCTTCGAGCTCGGA 14070  
Qy 1796 CGCTGCTGACAGACGCTTTCACCGAGCGCGCTGTTACCTTGCATATGCGCTCG 1855  
Db 14071 GCTTCTGACCGGACCGAGTTCAGCGCGCGCGCTTTCGCTTCAGAGCTGCGCT 14130  
Qy 1856 CGCGCTGAGCGCTGAGGAGTGTAGAGCGGAGTGTGCGCGCGGCTGAGCATGCGTG 1915  
Db 14131 TCGCTGCTGAGGCGTGGGAGCTGACGCGCGAGCTCTCTGAGGAGATCTCGCGCG 14190  
Qy 1916 AGCTGAGCTGCTGCTGCGCGCGCGCTTCTGCTTGCAGACGCGGCTGCTGCTG 1975  
Db 14191 AGCTGCGCGCGCGCGAGCTGCGCGCGCTTCTGCTGCTGCGCGAGCGCTGAGCTGCTG 14250  
Qy 1976 CTGCGCGCGCGCTGATGACAGGCTGCGCG-----CGCGCGCGCATGCTGAGTGC 2032  
Db 14251 CGCGCGCGCGCGCTTTCAGAGGCTGCGCGCGCGCGAGCGCGCGAGTGCAGGCTG 14310  
Qy 2033 AGGCGCGGAGCGAGTGGCTGCTGCGGCTGCGCGCGCGCGCGAGCGCTGCTGCTGCT 2092  
Db 14311 AGCGACGAGGCGAGGCTGCTGCGCGCTGCGCGCGCGCGAGCGCGCGCGCGAGTGC 14370  
Qy 2093 CGCGCTGACGCTTCGAGCAGAGTGCATTCGCGCGCGCGCGCGAGCACCGCTGATGCGA 2152  
Db 14371 CGCGATCAGCGCGCGCTTCGAGGATTCGCGCGCGAGCTGCGAGCGCGCGCTGCGCG 14430  
Qy 2153 TCGCGCGCGAGTGGCGCGCGCGCGCGCGAGACCAAGGCGCTCCAGCTTCGATGCTG 2212  
Db 14431 TCGCGCGCACTTTCAGAGCGCGCGCGCGAGAGCGCACCGCTTCGCTGAGCGCGCT 14490  
Qy 2213 TCCACTGACGCTATGCGCGCGAGTGCAGAGCGCTTGGGCGTGTGCGCGAGTGCAGT 2272  
Db 14491 TCCACTGACACATGAGACCGCATGCTGCGAGCGCTTCCGCGCTTCGAGAGGCTGCA 14550  
Qy 2273 GCTACCGCGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2332  
Db 14551 GCTTCATTCGCGCGCGAGTCCCATGCTGCTCAAGTCAAGGAGGCTTCGCGCGAGG 14610  
Qy 2333 AGGTAGCTTCGCGCGCGCTATTTGGGTCGCGCACGCGCGAGAGTGTGCTTCGCGAGT 2392  
Db 14611 AGATCGGCTGCGCGCATCTAGGCTGCGCACGCTGCGAGAGCGCTTCGCTGCGAGG 14670  
Qy 2393 GAGTGAAGCGCTGACCGCGCGGCTGCGGCGACCTTGTGAGAGTGCCTGAGATGAG 2452  
Db 14671 GCATCAATACGCTGAGAGCGAGCGCTTCTGCTTTCGAGCTTCGCGCGCGAGG 14730  
Qy 2453 CGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2506  
Db 14731 TCTTCTGCGCGCTGCGCGAGAGCGCTGCGAGCGCGAGCGCGAGGCTGCTGCTGCTGCT 14790



Qy	2507	CATGCGTGCAGCGCTGGGCGCTGACGAGCGCGGACGCGCTGACAGCGCGCTGAGCGGCGCTCTC	2566
Db	14791	CCGGCGCTCCCGACGCGCGCGCGCGACGCTCGACGCGCTCGACGCGCGCTCTCTCGTGCCTCC	14855
Qy	2567	GGGCGCTGCGTGCCTGTGCTCTTGGGCGGACCTTT--CCGCTCAGGGGGGCGGCGG	2623
Db	14851	ACGTGGCCCGACATCGGCTGCATTGGGCGGCGGCTTCTTCCGCGCGTTTCAGCCCCCAGACGG	14910
Qy	2624	TGCCGCTGCCACGTACCTTGGAGCGGACGAGCGCTACGTGGATTCGACAGAAAGCCGACG	2683
Db	14911	TGCGGCTCCGACCTTACGCGCTTCACGCGCACGCTTCTGCTGACGCG--CCCGCGG	14967
Qy	2684	ACGCGCGCGCTGGCGGACCGCGCTGCTCCGCGAGCGGCTCACGACGAGTTCAGAGAGAGG	2743
Db	14968	GAGCGGACGCGGACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGCGGCGCG	15022
Qy	2744	GCGCGCTGCGCGGCGGCGGACCGCGGCGACGCTCGGCTTCGACCATTCGCGCGCGCGAGAGCG	2803
Db	15028	TTGACGCGGGGGGACGTGGGCGCGCTCGGCGCGCACGCTGACACGCTCGGCGAGAGACACCG	15087
Qy	2804	GACCCCGGAGAAAGTGCAGAGCGCGCGCGCGGACCGCTCCGCTTCGCGCTCGAGATGAGTACG	2863
Db	15088	TTTCGCGCCCTTCGCGACGCTGTCCCGCGCGCTGTGCGGCTGCGCGCGCGCGCGGACGAC	15147
Qy	2864	CAGCGCTGCTTGATCATCCTCGCTGCTTGGGGTTCACGAGCGCGCGCGCGCTGTCTGAGCG	2923
Db	15148	GGACGCTGTGTCAGCGCGCTCCGCTACGCGCGCTGTGTGGAGACCGCGCTCATCTCGACCGCG	15207
Qy	2924	AGTTCGAGATGCGCGCTGCAGCGCGGCGGGGCTCAGCTTCAATGATGTCAGCTGCGCTGG	2983
Db	15208	CGCGCGACGTCGCGGGGACGTCGTCTCTCGTCACGCTCGCGGAGCGGAGACGCGCTCG	15267
Qy	2984	GCATGCGCCCGA 2996	
Db	15268	CGCGTGCAGCTCGA 15280	
RESULT 13			
AF319998/c			
LOCUS	AF319998	49736 bp	DNA linear BCT 28-MAY-2001
DEFINITION	Stigmatella aurantiaca myxalamid biosynthetic gene cluster,		
	complete sequence.		
ACCESSION	AF319998		
VERSION	AF319998.1		
KEYWORDS	AF319998.1 GI:14210834		
SOURCE			
ORGANISM	Stigmatella aurantiaca		
	Stigmatella aurantiaca		
	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;		
	Cyctobacteriinae; Cyctobacteraceae; Stigmatella.		
REFERENCE	1 (bases 1 to 49736)		
AUTHORS	Slalowski, B., Nordieik, G., Kunze, B., Blocker, H. and Mueller, R.		
TITLE	Novel features in a combined polyketide synthase/non-ribosomal		
JOURNAL	peptide synthetase: the myxalamid biosynthetic gene cluster of the		
MEDLINE	mycobacterium Stigmatella aurantiaca Sga15		
PUBMED	Chem. Biol. 8 (1), 59-69 (2001)		
REFERENCE	2110452		
AUTHORS	1182319		
TITLE	2 (bases 1 to 49736)		
JOURNAL	Slalowski, B., Nordieik, G., Blocker, H. and Mueller, R.		
MEDLINE	Direct Submission		
PUBMED	Submitted (07-NOV-2000) MK, GBF, Mascheroderweg 1, Braunschweig		
REFERENCE	38124, Germany		
AUTHORS	location/Qualifiers		
TITLE	1. .49736		
JOURNAL	/organism="Stigmatella aurantiaca"		
MEDLINE	/mol_type="genomic DNA"		
PUBMED	/db_xref="taxon:41"		
REFERENCE	complement(2. .1306)		
AUTHORS	/note="ORF1"		
JOURNAL	/codon_start=1		
MEDLINE	/transl_table=1		
PUBMED	/product="putative aspartokinase"		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			

gene' cds

protein\_id="AAK57182.1"  
/db\_xref="GI:14210835"  
/translation="MPEIRGARQAGRTSALALIVQYGGSTIGDPERIKVNAKRCICIA  
AOKGHADVNVASMSGETNRFLKVASQITERRPERQDVVAIVVAGSGTGLVALMIGI  
TQKRRVAFSLHQVRIPTTDSFVKARIKSIDARIIDALKNKALIVVAGSGTDEIGN  
VTTGKSGSDTTAVALAAAKADACESITYTDGVTITGDDPMNAKRLBEIAIEMELE  
LASGAKVIGTISRVEFMKRYKPLMWKSSFTDDEGTCEDENAMBEVLVSGIAYDKN  
EAKIANGVDPVPGVAKITFGALDEOSIIVDILIVONVSKGRGTDYTPYKADLTAK  
DVYKIKASVKAEBEVTDDOYKRSITVYGMGRHNSGVAAGKPTVLAGAGNVOMISTS  
EIKVCSYHSNTEYLAVRSHTAFGLDKPEAGVSEVSLNG"  
1363. .2895  
/note="ORF2"  
/codon\_start=1  
/transl\_table=11  
/product="putative histidine ammonium lyase"  
/protein\_id="AAK57183.1"  
/db\_xref="GI:14210836"  
/translation="MSRRPLNIDDDTLKLEELIQVARTVTELAAPAAVRAADL  
VDRAAGTPSPYIGTGRPTLAERVIDKDKLRELORNLISHAAGSPPLPEARVL  
LLKCNLTAKYSGISIRPETTALALEMNRVVPVPRGVGASGLDAPLALHAPI  
GEGEFTYGEGLVPAQALERVAGLKVLEAKESGLANVGNQACAVGTLQIAPAEHIA  
DLADLAGMTEGLGLSGHKPTEPEIDVRAHEGQKACAAHLRELADSLAVESHVNC  
KVOPYSKCPQVGHAGARBELSPARILVEVINSIDNPLVEVEREITYSGNVFNG  
FVSLADVAAALATOLSAISERRVEGLVNPLNSLPEPLANKSGNGSMFADGNTAA  
LVAASRVISHASVDSIPSSAGREDHNSMNTAKROVADTFRSLAELIIVAAOA  
LDYKQPTPAGKGPQAAVELIRSKIPTMEKORELHDIAVNSGALISDELLNARTATR  
GQ"  
complement(3275. .7822)  
/gene="mxaa"  
complement(3275. .7822)  
/gene="mxaa"  
/note="non-ribosomal peptide synthetase; contains a  
condensing, adenylation, and PCP reductase domains"  
/codon\_start=1  
/transl\_table=11  
/product="mxaa"  
/protein\_id="AAK57184.1"  
/db\_xref="GI:14210837"  
/translation="MSVRELLTTLAAQVRLREGDKLQVEASKAVITPELRBELSNAN  
KADILAFPKRKSSREETSSEICOPLARBEVYPLSSGGGERLMPDRUSPGTSQYVNHG  
LRAGALDTFKRLSDELVRHREVLRTHPPEVSGPROVYVAPDGVALTIVELPOL  
TESOREELIYRSDDELAAQOPEDLAKGLPALTVALGDPFALFVTKHMTITDGMSS  
VPESEELVYASPIRGQAAALPVSOMFADSLAERMAIAGESAGBARAKWKKLGL  
PPLQVDPHAAVISTSHRGATVPLASPSALSDALBELASBEGCSFTVLEAASALIH  
RYSQGVDPGVGTIVANGSVPADLIGFIANTLALRCDSPEPTFSQMLAAHNVLEA  
LDYQALPSEVSVQTVAGASRDGILNPVRACTFESIAPLTDLPGTSMSPLANAPDS  
VEGVAKPELTSILILASEKGLAGMLEYREVDVASTAYMKGQVLLSESIVAREPEV  
SKLPLTAEESEGRLINDMNGVLDVLPDCHNELVQAAKTPPOVAAVSOCKITLVAE  
LNRANQALAHILRLGTQKEERVELCYERREDIVIGLALIKAGAVAPDPAKPER  
LALIDEDGTVQLVLTQOALVPELPATQARVVCCLKDMPTIGASDTPETITPAEPIA  
YLITDSGQVCPKQVMIEHRNAVAFILMANSVSPKELACTLSTISICPLSVFELT  
PLCCGAKYAKNALDELPELPAAREVLTINTVPSAMGALRSGSVGSAIVNIAGASA  
LAGALVIDIYOLDHVRDVFNIGSEPTTYSIPFRVNGQPTIGRAGVNGTQVVALDS  
NRBEMPQIGREPVYIGGMGVARGVIGLPELJAEFRVSPVGGSEBARLYRTGDLARL  
PDGQELTEAGRNDQVYKLRGRIRBELAGEVIMESGIRDAVVVVYRBSGLDKQOLAVV  
VGRGEKAEPEERLHDVYKSKLPEYMPVPLTVGLDALPLTENGVDRLALPABRTSHG  
PAKEVHEPVTGEEESLAIWROYLVGEVQIAGHDFELGSHSLIKVILVASASAGA  
DIPRALIQATPLEEMARAVEAAKTGSLPHADYVVEAEADVALDAETALGKALPVTG  
ALRTILTGATGAGFLAELEICRGTARICYLVRSTBOEGMARIKKNLESYLNWAE  
ALAPRVVPDIDGICPLGLSEKPEOBSLESDIAIYNNGALNAPFVDEPRAVNYNG  
TREITRLATRRIRKPLHYSTVSVLPIGRKAPITREDEPLESPSLVGGIYQSKVNAEK  
LVRESRRGLGVTILRPGRVYGHSRTAADDJLVCCTLLGCVMGVAVSDVALDLDIT  
PVDVYSALIVLSMRPESIGQVYHNVNPOVRADDEMNVNARGVYGLRVLPYQOMISE  
LGSASBSSESLGDLMPFLQGVPPEDRSVGGPRMVVCDSDTLKALGCTGSCPSVDAS  
LISTYSSLVHRGFLKAPVYR"  
complement(7819. .11520)  
/gene="mxab1"  
complement(7819. .11520)  
/gene="mxab1"  
/note="PKS; ketosynthase; acyltransferase"  
/codon\_start=1  
/transl\_table=11  
/product="mxab1"

```
/protein_id="AAK57185.1"
/db_xref="GI:14210838"
/translation="MRAGAGGHPILGVSTTSVQKKNFWSALSKEBSYADHRVG
GAVVVGAAVYEMALSAARELSGEKAYELVDTSFKKGLLPQSGERVVOMALSGGCD
ALSPQSSQKASTEPNAPAGMAHVVQIRPIEGSPSTQTPSPFALIERCEVIAQA
AHYELASGKAVGPAFGVQOQIMRGRAALGVHLPBFLARASAVRIHPALILACF
OLVAASVPOEDLTPDAGPAVVALAHIRLHEDPGTEVPCHTVALRPSBGRVYFEDL
VLRDASGRILVEALRVOQLDPAAPAHAGKDLFFALPBRHASVQVQPAEPGASGR
WLLIADSELDATVESQLERHGEVIRVDLAGAGGRVVDATSPOADGVLSAFA
GASPRGVHLTGLEPREVDPIMGCGVHLTVQALRSVSLASPPRLMLVTRGVDSGS
RSASPVQASLWGLGRLTAHSEELKTRVDVASMGTDGAGALMLRELVAHDEEIS
LRADRGVYGRIVRSGPGRAPSEAVPAGRGPFHLERDAQLEFRAPARLPGGEL
ELVEVALGCGEGCGPERMEGAGEGICGSGRVVAGIGGVNVAEGERIALVAGSL
GSHVVPACCTAIRPSSSHGAAALAAVLPANVALPHLGLERKRWVIGATISGL
GSAIATKLARACAEVFAVATDEQPSARELDGADQVMDPSFVSRLDADTGGGVE
LAIVAPGAVELERSLPTLSEGRVLDLRASGASARTGDANVAMCVVDQDFARRPER
FASLQVQVAAFEAGGLPSAADARALAKDKARVVVPAADARLRADQTLVTGGL
GGLAVAKMVEGAGHVLVGRDVLVPEQOEVTVALAAGAVRVARADVSRDQ
IARITSEIAEGGAPLRGVTHAGVLDGGLTLOOTVBRPRVMAPVVIGGMNHLVTRP
ARIDPFVLYSSAAMFGAPGCGNYVAGNAPFDALAHHRALGPELSTIMGPFVGL
AAQANRGRRLQKSDSLTPAESNTILGRLDGDTQMAWVPLELRKVEYFPKAS
SPWSELVPADAGVEGEARDVALLETLTATPOBAKALEFVEEQULARLUDSAR
IDPEAPLOSFGILSLMGLERNRLVSGLSGLVSLPASIWKHPITLALCAHLGEVMDRT
LAETLQAAGAGPAVDESAADKEVFVL"
complement(11487..14303)
/gene="mxab2"
/complement(11487..14303)
/feature="PKS; dehydratase; enoylreductase; ketoreductase;
ACP"
/codon_start=1
/translation="MSKOEKPAEPSPALORALALVEKMQARLDAVERATEPIAVIGL
GCRPFGAVDGSYWKILRDGDVALREVDESDVDPGHDSRDVGRKMYTEFAGFLD
DYEHDAEFPGISPREASLDPQOQLVLEVAEALBNAGLADQLAGSKTVPMKVS
SDYMRLEKENDATRFDMATGNGSFVPGVSTYLQAGCKMFDVADCSVALH
LASERLRAGESNIALAGVNLILSPETMLCISQALASDGCKFDSADSVYEGE
CGVLKRLISDAQRDGDILALIRGSAVHMDGASGLTVPNPSSQAAVQORALDNARI
APALVYIEAHGTGTPLGDPLELRAGVILGKRPADRPFFIGSVKTNIGLEPPAIGI
AGVITITLISIOKEIIPHLHPRTNPVHEDWRIIPARVPERVPAHAGSRIAGVSEF
GLSINAHVLEAEASTAPTPVAEDKQQLVLSAKTPRALSLAEMBSRLKMBEG
GSLADISTYSALSRTPHEHRLALAKSKOELIQLAEAPAGSHAGLSGSGRTGLQO
KIVFVPGQSGQMLGMQRLVAENVTFRALIDWCHALISRTDMLRLDLTNPQOPI
NINDIVIDVLPALQVLAALVVRMLGLEPHAVHSGMSVLAHAGALSIEDARI
CERSKLRRVSGQGAAMVVDLSRBOAEVIRGAEIRIALAVNGRSVTLSGDPALK
EVLRLERREVCRWVKVDVASHSPOMLRQDLDRMAAIRPTGSAVDLYSTVGAP
IDGRRLSAPYWRNLROPVLPADVQRLQDENVLPILSPHPIIIVPVDMALREGDT
SGRGLVPSLIRDOEWPSSLSSVQGLFTFVIAWTGSGITPTSVAGLRFRLTRGSASG
TGMSCQQLRRRGSCALARVIPS"
complement(14300..29878)
/gene="mxac"
/complement(14300..29878)
/feature="PKS; module 1,2 and 3 contain KS, AT, DH, KR, and
ACP regions"
/codon_start=1
/translation="MANRDESNELSAPAKMLVALEKQTRLNAVEGAKEPIAVGM
ACRFPQADNVFAFMRLQGVDAVREIPIERMPKQEVLDIGGARNVFGALDVSVPD
FGFPGISPREAVFMRDPOQLLEVAEMALDGLVDKLSGASVFCGACDDVHFD
OYERBETGAFSAGTAVSVLSRGLSVYENLQPSLVYDNTAGSSVLTAHCOGLRN
BCNMALAGCVNLILSPOSVILSVKQLASDPGRKAPDASGSTRGEGCGVILKX
ISDALADGHIILATTRGSHINDGKSTGTTTNNVSQLITQALSSGLKAEQVSYV
LAHGTGTPLGDPLEAEALRETYGVPRDPSVCGISVKTNVGHLSAAGIATIMVVL
AMRQSIIPHLIKQVNPRIQLEGSALTITPRLTPWKSAGQPRAGVSGFSGISNNAH
VIEBAPAVAKDALPQREPVLPASRSSEVALAERVAELGREGGSLSDCYTA
SAHRTHTHRLVLADSRHTMERLRAFAFRGBOGPGASGRAPFGRRRTIVVYVYGG
```

```

AORAGMBALMESEVPFDALVKVDKALQPLHGLMSVLDEIADREORSRLAIDVNOPV
LPAYVALTALMRSMGIEPDAVMHSMKEVYAAHVAAGLSIEDVLIITCHRSKMRTL
GKRGATAMNVLSEBESAEIRPNMGKIULGLOGPREQVYSGEBAYKEAVEALGKR
GVFRVWMDVPSHTPLVEVTSRELEGIATITPRAQIPLCSTVTGFLGAKIDAA
HMGNLRNRPFAAQAQHLASDGTPEVSPHVLPAVQITMSAGTEVLAISLR
RNEBERHTMLSLSALYSSGRVWNNRYIPKQGLVLPPTVPMQEREMFMAVGR
ASRTSGSHPLGERIILDSAGEVRLMOQKPHFPLAHRVQGVALLPAAADAM
ALSAKQITGEBVLEIDVFSFALLSPDEDERVQIMLSTPGRASRLPAPMRE
TFSDBLTALHABESLRHGDASVYKAPVALDELBARCEBSEAYEATQTLKEIGLE
YGPSPRVVRLRKLKGEALGVQLEGTAAEERSLHPALLDCFOVLAALBGLDS
EGGVVPAVLSRLVLVPEALMSHRIVSGAEGSNTIECDVTLTLDQORVAEVR
RLVCKIILAAAPADEVDNRVYSAQMEPOQQALETSSRTAGRMVLFQAPARIGQS
LAGIAERAGCVVVEAGVSYERLEADRYINPSGLLEDFOLALKEAFQDQSKCMVH
LXGLDVANVTPSPSPFALLSSTSVYILOAAQORMDTPMLMVTGTOVYASG
EQGSPAAAPLVGLAAVTHHPELRCHVDLSAQPSPEIASLTIELLGGABERLA
LRGAVRHVARLRGKISGKAAEPREFADVLYLTGGJGGLVLEPAKFFIORGANILA
LTGGGAPSEERQKTVLELRNAGAEVRVRAVNTSAADVARLGSIDSLPLRGVHA
AGVIALDGLVLTDLDEPRLAVLAPVQGMNHLVOTQGRPLFLFVMSAASMGSPGO
GNVAAANFMVULAHHRRAQGLPGLSVMGFMADVGLAAVATNGDRLALRGVGMPP
AQLAEALRMLVQBAQVAVVPLDLRQWREERYLSAASPFUSVLMQKRVSTPASRKS
ARELLAAAEPRQKILEGYRBDQGRILRNEPRRIEDQPFQDGLHMSLTGLBLNR

Query Match      19.8%; Score 844.4; DB 1; Length 49736;
Best Local Similarity 58.4%; Pred. No. 2.9e-83;
Matches 1563; Conservative 0; Mismatches 1091; Indels 24; Gaps 4;

1 GTGGCGGATTCCTCCATTCGACGCGCCGACCGCAAGATTCGGATTGCTGCGGACCGAT
19792 GAGGAGGAGACCGCGTGTGGGCAACCCACGAGCGGATTCGATGTGCGCATTTGGC 19793
61 TGCCGTGCGCCGGTGGCGGTATCGATTCGACGCGGATTCGGAACGCTCTCGAGGCGCTCG
19792 TGCCGATGCCGAGGCTTCG---GCTCCGAGGCGCTTCTGGAACGCTTTGGAGGAGAGC 19676
121 CGCGACACCGTTCGGCGAGTCCCGCGCAACGCTGGATGACGACGCTGTTGATCCC 180
19675 GTGACGCGCATCCGGAAGTGCCTGCTCGCGGTGAAGTTCGAGACCTACTACGACCCG 19616
181 GACCCGATGCCCGCGGAAAGACCGCGTTACGCGCGCATCTTTCTTACGACGACTAGCC 240
19615 CAGCCCGCGTCCCGGGAAGACTCAACCCGCTGGGTGGTTCCTTGGATGGGCGTAC 19556
241 TGCTTGAAGCGCTCTTCTTGGGCACTCGCGCTGCGAAGCGCTGCGGATGAGCCCTGCA 300
19555 CAGTTTGAAGCGCCAGTTCTTGGGATCCCGCTGCGAAGCGCGCAATGAGACCCGAG 19496
301 CATCGACTCTTCTGAGAGTGTGCTGGAGCGCTGAGAAACGCGCGCATGCTTCATCG 360
19495 CAGCGCTCTGCTGAGAGTGTGCTGGAGTCCCTGAGAGACGCGGGAATGCTCGCGG 19436
361 GCGCTCTGCTGAGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTGCTGGAGTGTG 420
19435 CGCCTGGCAACACCCGCACTGGCGTGTCTGATGTGATGAGCAAGACGATGACCGATG 19376
421 GCGCTGCGCAAGCGACGCGCTGCGCAGAGATCGACGCTCACTGCGCGGCTGGGAGCATG 480
19375 CTCMAAGGAGTGGGAGCGCTGCGCGCAGAGGATGCTTCACTGCGCAGCGGCAATGAC 19316
481 CCCAGCGTGGAGCGCGGCGCAATCTGTATGCTCTGCGGCTGCGAGGCGCTGTGTCCG 540
19315 TCGACCTTTCGCGCGCGGCTGCGCTATGCTGCTCGGTTTCAGGGGCGCAAGATGAGC 19256
541 GTGATAGCGCTATTCCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
19255 TTGAACACCGGTGCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 19196
601 TCCGGGAATGCTTCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
19195 GCCGTGAGAGCACTGCGCGCTGCGCGGTGCGCTGCAATGACGCTGCGCGGCACTCG 19136
661 CTCGTGTGCTCTCGAAGACCCGGCGCTGCGCAGGACGCTGCTGCTGCAAGGCAATTTTCG 720
19135 ACCATCTACCTCGGACGCTTGTGCTGCTGAGATGGCGGCTGCAAGACCTTGCAT 19076
```

Qy 721 GCGAGGCGGATGGGTTTCGACGAGGCGAAGGGTGGCCGTCGTCGTCCTTCAAGCGGCTC 780  
Db 19075 GCTCCCGGAGCGGCTACGTCGCGAGCGAGGGGTGGCGCATCGTGGTTCTCAAGCGGCTG 19016  
Qy 781 AGTGAACCGCGCGGCGGACGCGGATTCGGAATATTGGCGGATTCGAGATCCGCGCATCAAT 840  
Db 19015 TCCCATCGAAACGGGATGGGGATGACGTCCTGGCCGTCATCCGCGGCTCGCGCATCAAC 18956  
Qy 841 CACGACGATGCGAGCAGCGGCTCTGACCGTGCAGAACGGGAGCTCCAAAGAAATGCGTGTG 900  
Db 18955 CACGATGATCCGAGCAGTGGCTCTACGAGTGCCTCAATGCGCATGCGCCACAGAGTTCTC 18896  
Qy 901 AAACGGGCGCTTGGCGGAGCGAGCGTGCAGCGCTTTCGATGGGTTTATGTCAGAGCAC 960  
Db 18895 CGGCGGCGCTTGGCGGAGCGAGCGTGCAGCGCTTTCGATGGGTTTATGTCAGAGCAC 18836  
Qy 961 GGCACGGGCGACGACGCTTGGTGAACCCATCGAAATCCAGCTCTGAATGCGGTATAGGCG 1020  
Db 18835 GGAACGGGCGACGACGCTTGGTGAACCCATCGAAATCCAGCTCTGAATGCGGTATAGGCG 18776  
Qy 1021 CTGGGCGGAGATGTCGACGCGCGCTGCTGATCGGGTGGTGAAGACCACTTGGGCGAT 1080  
Db 18775 CAGGAGCGTACGTCGGAACAGCGGCTGCTGATCGGCTGGTGAAGACCACTTGGTAC 18716  
Qy 1081 CTTGATATGCGTGGGAGTCACTGGGCTGCTGAAGTCTGCTTGTCCCTTCAACAGCGG 1140  
Db 18715 CTGAGGAGCGGCGGCGGATCGCGGCTCATCAAGGTCGTCCTGGCCATCGGCGATGGG 18656  
Qy 1141 CAGATTCTGCGCACCTTCACGCGGAGCGCTGAACCCCGGATCTCATGGGATGATCTT 1200  
Db 18655 GTGGTCCGCGCACCTTCACGCGGAGCGCTGAACCCCGGATCTCATGGGATGATCTT 18596  
Qy 1201 CGGCTGACCGTCAAGCGCGCGCGGACACCGTGGCGGAGCTGAATAGCGCGGAGCGG 1260  
Db 18595 CGGCTGACCGTCAAGCGCGCGCGGACACCGTGGCGGAGCTGAATAGCGCGGAGCGG 18536  
Qy 1261 GGGGTGAGCTGCTTCCGATGAGCGGAGCAACGCGACGTCGTGCTGAAGAGCGCGG 1320  
Db 18535 GGCATCAGCGCTTCTTGGCTTGAAGCGGACCAATGGGACGTCGTGCTGAAGAGCGCGG 18476  
Qy 1321 GCGGCGACGTCACACGCGCGCGCGCGGAGCGACCGGAGAGCTGTGCTGTCCGCA 1380  
Db 18475 GCTCCCGAGCTCCGCGGAGCAAGCGGAGCGGAGCTGTGACGTCGTGCTGTCCGCG 18416  
Qy 1381 AGGACCGGTCAGCGCTGAGATGCAACGCGCGCGCGGCTGCGGACCATGTGAAGACCTAC 1440  
Db 18415 AAGACGAAAGGTGCTCAACGCACTCGCGGCTCGGATGAGTACCTGCGGAGCG 18356  
Qy 1441 CTTTCGAGTGTCTGGGAGATGTGCGCTTCACTGCGGAGCAAGCGCGGAGCGATGAG 1500  
Db 18355 CCAAGGAGAGAGCTTCGCGGATGTCTGCCATCGCGGAGCAAGCGGAGCGATGAGT 18296  
Qy 1501 CACCGGCTCGCGGTGGCGGACGTCGAGGAGGGGCTCGGCGACCCCTGAGCGCTGCG 1560  
Db 18295 CATGACCTCGCGCTGTCGTGAGAGTGGTGAAGCAAGCCGCGAGCGGCTGTGGCGTTC 18236  
Qy 1561 GCGGAGGAGAGAGTGGCGCGGCTGCGGAGTATCGGCGATCTCTACCGGCGAG 1620  
Db 18235 GCGGAGGAGAGAGTGGCGCGGCTGCGGAGTATCGGCGATCTCTACCGGCGAG 18176  
Qy 1621 CTGCGCTTCTCTTCAACGAGAGGGGCGAGACGTCGAGCATGCGGCTGAGCTGAC 1680  
Db 18175 GTGCTTCTCTTCAACGAGAGGGGCGAGACGTCGAGCATGCGGCTGAGCTGAC 18116  
Qy 1681 GATGTATGTCGCGGCTTCGCGAGCGCTTCACTGTGCGTGAAGGCTGTTCACCGAGAG 1740  
Db 18115 GAGACCGAGCTGCTTCTCGCGCGCTTGGATCGTGCATATGATGAGAGGCGCAAG 18056  
Qy 1741 CTGACCGGCGGCTTCGCGAGAGTATGTCGCGCAAGCCGCGCAAGTGTGAAGCGCGCTG 1800  
Db 18055 TGGGCGAGTCTCTCTCTCGGTATGC-----ACGGGACGGGCTCGCGG 18011

Qy 1801 CTGACCAAGACAGCTTACCCAGCGCGCTGTTACCTTCAATATGCGCTGCGCGG 1860  
Db 18010 ATTGACCAACCGAGTACACCGAGCCCGGCTTGTTCGCTTGGAGTACGCGCTTGGCGAG 17951  
Qy 1861 CTGTGCGGTCGTGGGGTGTAGAGCCGGAGTGGTTCGCGGCACTACATCGGAGCTG 1920  
Db 17950 CTGTGCGGTCGTGGGGTGTAGAGCCCGGAGTGGTTCGCGGCACTACATCGGAGCTG 17891  
Qy 1921 GTGCGCTCTGCGGTCGCGGCGGCTGTTCTCGCTTGAAGACCGCGGTTCTGTCGTCG 1980  
Db 17890 GTGCGCGCTCTGCGGTCGCGGCGGCTGTTCTCGCTTGAAGACCGCGGTTCTGTCGTCG 17831  
Qy 1981 CGCGGCGCTTATGACGAGCGCTGCGCGCGCGGCGGAGTGTGTCATGAGCGCGCG 2040  
Db 17830 CGTGCAGGCTCATGCAAGGACCTGCGCGCAATGCGAGATGAGGCGGCTTCCGCGAGT 17771  
Qy 2041 GAGGCGATGTGCTGTGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 2100  
Db 17770 GAGGCGGAGTGCAGGAATGCTGCTCTTATGCGGATCGGCTTGTGATGCGGCGGCTC 17711  
Qy 2101 AACGCTCGGAGACGAGTGTATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTC 2160  
Db 17710 AATGCGCGGCGGAGACCGGATGTTCCGCTGAGAGGAGCGGCTGAGACCATCTCGGCT 17651  
Qy 2161 GCGATGCGCGCGCGCGGCGGCGGCGGAGCCGCTCCAGCTTTCGATGCGGTTCCACTCA 2220  
Db 17650 TCGTTGAGGCGCAAGGCGGCTCAAGAGCCGCGGCGGCTTACCGTGTCCAGCGGTTCCACTCT 17591  
Qy 2221 CCGCTCATGCGCGCGGAGTGTGAGAGCGTTTGGGCGTGTGCGGCGGCGGCGGCGGCGGCT 2280  
Db 17590 CCGCTCATGAGCCCGAGTGTGAGAGCGTTTGGGCGTGTGCGGCGGCGGCGGCGGCGGCT 17531  
Qy 2281 CGGCGGTCGATTCGCTGTCGACGATCTGAGCGGAGAGGCTTTCACAGAG---CGAGGTC 2337  
Db 17530 GCGGCTCGCTTCCAGCTGCGCTTCCAGCTGAGCGGCGGCGGCGGCGGCGGCGGCTG 17471  
Qy 2338 AGCTCGCGGCGGCTTATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 2397  
Db 17470 CTGACCGCGGAGTCTTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 17411  
Qy 2398 AAGGCGTGCACGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 2457  
Db 17410 AAGGCGATTTGGGCGCTGCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 17351  
Qy 2458 CTGCGGCTGAGTCTGCTGCTGAGTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCTG 2517  
Db 17350 ACGAGCATGCGGCTGAGGAGGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 17291  
Qy 2518 GCTGGGCGTGAACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 2577  
Db 17290 AAGGCTCAGCGCGGAGCGGCGGAGTCTTTCAGCGCTGCGGCGGCGGCGGCGGCGGCTG 17231  
Qy 2578 GGCCTGTCTCTGCGGCGGCGGCTTTCCTTCCGAGG---GGGCGGCGGCGGCGGCTG 2634  
Db 17230 GTGCGGCTGAGATGAGTGTGAGTGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCTG 17171  
Qy 2635 ACGTACCTTGGAGCGGCGGCGGCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2672  
Db 17170 GTCTACCGGTGAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 17133

## RESULT 14

SAU421825

LOCUS

DEFINITION

KEYWORDS

ACCESSION

VERSION

KEYWORDS

SAU421825 56808 bp DNA linear BCT 16-MAR-2002  
Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,  
ORF1, stia gene, stib gene, stic gene, stid gene, stie gene, stif  
gene, stig gene, stih gene, stii gene, stik gene, stil gene and  
ORF9.  
A421825 1 GI:19572309  
acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;  
kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;  
ORF7; ORF8; ORF9; polynucleotide adenylyltransferase; potabium

SOURCE	channel beta chain; ribosome binding factor; stia gene; Stia protein; stib gene; Stib protein; stic gene; Stic protein; stid gene; Stid protein; stih gene; Stih protein; stif gene; Stif protein; stig gene; Stig protein; stih gene; Stih protein; stij gene; Stij protein; stik gene; stil gene.
ORGANISM	Stigmarella aurantiaca
REFERENCE	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Cyctobacterineae; Cyctobacteraceae; Stigmarella.
AUTHORS	1 Galatzis,N., Sliakowski,B., Kunze,B., Nordstiek,G., Blocker,H., Hofte,G. and Muller.R.
TITLE	The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmacellin is directed by a novel type of modular polyketide synthase
JOURNAL	Online Publication
REMARK	J. Biol. Chem., 10.1074/jbc.M11738200
AUTHORS	2 (bases 1 to 66808)
TITLE	Muller.R.
JOURNAL	Direct Submission
REMARK	Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124 Braunschweig, GERMANY
FEATURES	Location/Qualifiers
source	1.. 66808
	/organism="Stigmarella aurantiaca"
	/mol_type="genomic DNA"
	/strain="Sg a15"
	/db_xref="taxon:41"
	complement(4978..6231)
	/note="ORF8"
	/codon_start=1
	/transl_table=11
	/product="cytochrome P450 monooxygenase"
	/protein_id="CAD19077.1"
	/db_xref="GI:19572310"
	/translation="MLRSVSRPPSPSPANPVAAPFLARLDAFYADPPFLYAMBERD PHEWPDGSLFLTRWADLCYRDRTRSDSDRARGAKYGDAPLFEHHTTSLVNDP PLHTRVRLITVALTPRALSTWERGLRTLVLLGLAVKAGADIEDPAALPIEVEI GULDIPTREERPLRGWSLAIIGALEPRLTAQECGNNAVTEFLDYLRKILVAORAR RQDPATVDVTRLIQESDGERLTELHQCVFLNAGHETTNLIGALELAFPPD ERARILRAPALIPTVAEEVLRYESSNQLNRRVAEDTEIGVAVPTGTFELTICGAN RPARPEDEHEFDVGEOPRHLAPAGGANTCAGMULARIEARIALAFLARPDVALT APVARARARFPGFTAMPTRLGSLR"
	6369..7730
	/note="ORF7"
	/codon_start=1
	/transl_table=11
	/product="protein kinase"
	/protein_id="CAD19078.1"
	/db_xref="GI:19572311"
	/translation="MBELISEALPDGSRVGPRIIGRAGYGTGAVTRAHVYRICPR SVALKLARYPNDLRFEREARELLERVAHANVPRLLGHGWRAPOGEAHYVVMQWEG VLYIQMKERLTAREVLVLAQVAVARLALATMRGQLHRDVGENVLSPESGAILMD FCGTEGWAALPLTGLIAPGTAKVSRPEALRPHMNRNDAADAHYQATPDVYALGVT AVRLCTGIYPATPSGSGEDERETQEKRPQSVORPLVPALBDLILMBLNDKCR GRAGIELAEMELABRAGAEADBPNNRHSIPSEITGFPVQVQEMGMPALVLYVA VLLLMGSSGLNTRRVAPASRWTGSGTGVADVAVEALPVSSAREPSPGLNDML KPLPGQRRPPCPRFQISIQGGMLEVPQSSPPCGESYYEMKACYSVPAAPRRTS DPQ"
	complement(7849..8577)
	/note="ORF6"
	/codon_start=1
	/transl_table=11
	/product="methyl transferase"
	/protein_id="CAD19079.1"
	/db_xref="GI:19572312"
	/translation="MSQNTYDDSPESYSOLARSLKGLAGAPEMPALRALIPVXDL RYVDDCGRGRCRNARKNDAAHYIGLDSSEKMLRAKETSGDAITTYQADLEMLV PGAFDLAFSSIAIHVELEPLATVHGLTPGGFVFSIEHPIMASROPEMNGTP EQRHTWPVDSYOLEGRFTTHWLAEEVVKYHRTLGLTNALLDQGTLLRIIEBGMPTDA QLSAHELAEEBDRDPMFLVLSAQR"
	complement(8622..8879)
	/note="ORF5"
	/codon_start=1
	/transl_table=11
	/product="acyl-CoA binding protein"
	/protein_id="CAD19080.1"
	/db_xref="GI:19572313"
	/translation="MALEDPRSQERKALKTBPNDRLLETLSLFGKQATEGVQK PGILDYERAKYAWYGRKGLGRSEMQVVALVERLNRG"
	8961..10070
	/note="ORF4"
	/codon_start=1
	/transl_table=11
	/product="potassium channel beta chain"
	/protein_id="CAD19081.1"
	/db_xref="GI:19572314"
	/translation="WVYLCSRQSLRWITLNGSAHRRFRAMGCHARRPRMFRHLGRS GLWSEISYGMWLTHGSOVEEAAVACRAALDGGITFTPDADYAGTAREVIGRAL QGOEREGLEIPTKYWPTGPNDRSGSRKHLISIGSLRLKTEVVDLYQARVDT ETPLEETQAFSDVROGKAHYIGSEWTADQLREGVLSROMGFOLVSQPOYSLMW RVIEAEVPAQSEIGVGQVYVSPVIAQGLTGKVRPGKPPGSGRATDEKGSFTRP LNDVLTFRVQDLTPVAQEGSLAQLAQVAVYLNQPNNSALITGASREPOQVENYKAG VLSBEARHIDTLGLPTVERNPKTLENSPRKP"
	10428..10862
	/note="ORF3"
	/codon_start=1
	/transl_table=11
	/product="ribosome binding factor"
	/protein_id="CAD19082.1"
	/db_xref="GI:19572315"
	/translation="MTSEFKQRRRRAPRGTESSLSFSAQSLPAHHLRVOSTLFOEVA RLFGELSDPLLEEGSVTSFELTPDGNARIGFTLPPEYAGSGAPTEALVRASGFI RAQLALGNLKRVRHLRVCVAPAPGSGIPEDSAEKGSBS"
	10859..12016
	/note="ORF2"
	/codon_start=1
	/transl_table=11
	/product="cellulase"
	/protein_id="CAD19083.1"
	/db_xref="GI:19572316"
	/translation="WRPRRIEAGANGVPIIIMARAVPAGAVKQOLAROPYLAENVA AMPDVHAGVAGVGFATERHIVPGALGDLGCVSARVDFPMSIGRVELTRLLA APAAVAVPDGVAHGRGSLPPGLDIGLSTRLOREWERLVPHLGLTGSGNHLEL DQDEBGAAMLILHTGSRGVGAIAAHHLRVAVQSGESLPOLRDTSGAACLADTEL ACOPARANDTLORAVEVLAGAIGQGFVLDRAVDVHNHVAASHSFGRTLLVIRKA GVIAVGOQGLIPGSMGTASVYVAKGGEQARCSHGGRVTRREARERIRPALAH ALRVYDPBGRVTLVERAPTVARVDITEVLEDEABLVTPRLRPIANLKG"
	complement(12082..13974)
	/note="ORF1"
	/codon_start=1
	/transl_table=11
	/product="cellulase"
	/protein_id="CAD19084.1"
	/db_xref="GI:19572317"
	/translation="WKSLLILMLAVLPIYTAALARPANABEGHFEGBENAGLTA AGCPRPVLESTRGYIVDRCEGRKLSVNNFGASDLEVVGGDKOQLSDLVAGMKA MGPNSVRLPFPENMLHPDKNPAORCLEKTSVTCIDSTKNPELINTKALVSVAGMA ELTROGLVLIINHTTKSMCGMDGNGFMSQALODWODMTLAKRYQGNKVAV ADLAEVNPDDGLSDSPGMRQMHMAAOMGLMLRTNDDLIVERAVMNGGLDQ SRPOLKPRQRPVALLRDKLYVYAHVNTGYTRPNQSGSLSSGKPYSDMDCTYXGL DQEMGFALAAQATAPAWKSEFGVINEQVANSRANFSLADLVLEADVMAVYALN AAKLQNSVQGEDETYGILMSYPDMNGVRSGRLTGPRLGLADGRTGAVDAVRFPL MTVLLKDGSSYYVDNNSLSDMSGKAKISCPBGYRWGVSAFTILCTNAGVPAQ OGTGSYORTITQEVSPRLRYGDMANOSIKFECPTGYAVGISANPFLLEAGLCEON TGCTPINKSAKNFMWGDQRASSSGGMSVSQYKQCADNXYLILGLARRSGLDVPS VVLCSN"
	14345..21466
	/gene="stia"
	14345..21466
	/gene="stia"
	/function="pkgs"
	/codon_start=1
	/transl_table=11
	/product="Stia protein"
	/protein_id="CAD19085.1"
	/db_xref="GI:19572318"

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

gene  
CDS

/translaction="WTOLASRLGVGSESIDPRERSRYGIDSRDTKSLIAELGALLKR  
PLSLMTATYPTVEALSRLHAGEADVSGERADTLGVODDEPIALIGAGGEPAPAD  
BSFSLMTADVDATTEVPERMDLAFYFVKGKSGKNTMGMDLVDVDEDFLPGF  
ISPREAIOMDPOORLALSWALESASAVADSLMDSTGVMGAMGMDVAKLRLGAL  
GLIGPHRTAGDSSIIARVSYFLGRPSLTIVNRCASALVHMLACQSIINBEAAM  
ALAGVNLISPESSVAMSQJALSPORCAPFDRNAGVYRGGGGVVLTPSLRAI  
ANGVPLVCIVIGSVAVNDGFSNGLTAPNPAQOEVLRAVAVAGSPDKVHYETHGT  
GTIVGDIPIEVALASVLGNVPAERPLIGSVKTVIIGLEAKAGIAGIMKTVLSOHR  
AIPLNHEOPNPYIRPEELRLVKSLHPNNEGESLADGSSFGFTGCHVLEA  
PPSASAHIAPIAETPEELSLARQARAVVSGVADLCEPVARCEGSGORLA  
LIASPAALAAQVDFLAGERKPLIRBEVTSQSQKVPFPGSGSVPWPMGBELL  
TEVPFGAIEKDRALRGLSGMSMLEEISAGBASRLKESGLQALPILAVGLAAM  
RWSGIEPVGVSIGEVAAHVAGIIMLEDAARVIMHLSMGLMSQGMIVTDS  
AEANRLIREWERHVVVACNAPASTVSGAREVIEELGAELEGGIYANHRTIDPS  
HSLPMDPMGLAEVLAGISPRVAVPWSVTGTGFLAEECTGTAWYRNAREPLDPS  
OALTKLFEODRLPLEISPHPLIARVSOCCGREGROGVARSIRGSEBACRESL  
GRUFTTGSVAMQALYPAEAKRVELPMELGTGSLDLELRGARLLPLSAPHTAELEDA  
ARALQASTAKHPDRLDLCYSASVGSVHARILAVTDSKRALVEHLDDFLKESRP  
GRIGAVCATGEPKRVFVFPQSGSQMGQOLRQEPVEFEALIEECALIRSCAGWS  
LIDALQADEPHSRIGSIEVIOPTLPAIEVALAALMRSGVMDAVGHSMEETAAV  
AGALRTDAARVTCERSQLRRVAGOGGMALVELSELBOAREMLSKDGRVSIAGNSP  
RSTVSGDALIEELRELEAOCTGCRVKNVAVSHSPONDILRDLRISTAPRA  
ATVPLXSTVTGEPLEGAMDSYMMRNIREVYLRQALERLAGGCHGIVLSHPIL  
LPSIDQTLRHSKGVLPRLREPERAVLLESGLTLFTGGGRVDFRGLHPLGSHRV  
RLPHLPWQRQFWMIDMDVAAAGASVGAAMPGGSEHPLGHRPLSLAHAPCHVME  
NKLRSASVSLDIDHVGCTVLPSTYTEMALAAKSGKTVLSLRVLEBKMSP  
ETASRTVOLSLIOEBAGHSFRLHGRSPDEPTWLLGGGTAVHEGDVPPADPLMA  
LAVETVLOGKEVASDIAVEAMHAGNEKXPLPRGIEOMVWDAELVALRDEPLEOE  
LPEYRHPAALDLCIALIVAAHCEGGLAPTRLESIRAHASGAMKCHVNLNPV  
REASFSGEVLPDESQORLAEQARQILIGSRPHDGLVDMLEHVMQPLPV  
QVPEVRKLFGLASRSTTILRGIVHLMELDAPSPREGVNDLSRRLRSGSVSLIIG  
GLNAAELAOQLMLVTRGVQVAGSAGGAAVQSPFMGGOVVAAYHARLCRTAVD  
LAPVSPREEOTLLOELIAPAGDOVALRGRBHRPLTRCTREBVAERSDLEAGAI  
RSDCYLITGLGIGLTIKAKVYAKGRARHIVYGRRASDAEIALCOMGAGAECL  
VMKADVANTSVEELVGTIQRKMPPLKRIFFAAGLDDGTVALQOLARVVMKQVE  
GAMNLRATLASLDFVLFSSASLVIPOGNYAANAFLDQLAGLRKARGACIOS  
IOMGPSEIIGLAEBQDNRGERLALRGMGLNPREBELMLLSGEBLRRARLIGV  
RLDARRWCKSPAIRSSPLFSVLLAGEENAOVERLEGRLEARNDEPRTARLESF  
LREAAOYLRLPPARIDLRSLOTLGMSLSLLEARNLENTLETLDPATVLYMPTL  
SALAPFLAERMGIOANGVEBPQAVESVSEBDEGAMLRLEORQELSNEDLGMLS  
DKKEDEVSS"  
21463..26220  
/gene="sc1B"

Query Match 19.2% Score 817.2; DB 1; Length 66808;  
Best Local Similarity 57.5%; Pred. No. 2,4e-80;  
Matches 1597; Conservative 0; Mismatches 1148; Indels 31; Gaps 6;

21 GCGGCGACCGAAGATCCGATTGCGATGTCGAGACGAGTTGCGCTGCGCGGTGCGGT 80  
DB 43614 GCGTCTGGCCCCCGACCCCATGCGATCTGTGGCATGCGCTCTGCGCTACCCCGGTGGGGT 43673  
QY 81 GATGATCTGAGCGGGGTTCTGACGCTCTCTGAGGGCTCTGCGCGACACCTGTGGCGAGT 140  
DB 43674 GAAGGACTCCCACTCTATTGGGATTTGTGAAGCGCGCGGAGATGCGCTGACAGGAT 43733  
QY 141 CCGCGCGAAGCTGGGATGACGAGCGGTGTTGATCCGACCCCGGATGCCCGCGGAA 200  
DB 43734 TCCAGCACCCGCGTGGACATCAACGCTTTCATCATCGAAGCAGGACAGCCCGGAA 43793  
QY 201 GAGCGCCGTTACGCGCGCATCTTTCCTGAGCAGCATGACCTGCTTTCAGCGCTCTCTTCTT 260  
DB 43794 GACGACGACGCGCTGGGGGGGCTTCTGATGATGATGAGGGGTTGATCCCGACCTGT 43853  
QY 261 CCGCATCTTCCCTGCGGAGCGCTGCGGATGAGACCTTGCAATGATCTTCTGAGAGT 320  
DB 43854 CCGCATCTTCCCTGCGGAGCGCGGATGAGATCCCAACGCGCTCTTCTGAGAGT 43913  
QY 321 GTGCTGGAGGCGCTGAGGACCGCGCATGCTCTCATCGGCGCTGCTGCTGATGAGGAAAC 380  
DB 43914 CGCCTGGAGGCGCTGAGGACCGCGCAACCGCATGCTGCTGCGCGGAGGCAAGAC 43973

QY 381 GGAAGTTCATCGGAGATCGGCGCCCTCGCAATATAGGCGCGCTGCCGCAAGCAGCGC 440  
DB 43974 GGGGCTTCTGCTCGGCTTTACAGAGGACTTACTCTCTGTGTC-----TCTCGGA 44027  
QY 441 GTCCGAGAGATTCAGCTCATGCGCGGCTGGGACCATGCCACGCTTCGAGAGCGGCGC 500  
DB 44028 TCTCAAGCCCATCAACGCTTATACCGGACAGGACCGGCAACACGTTGGCGGCTGAGCG 44087  
QY 501 AATCTGATGCGCTCGGCGCTCGGAGGCGCTGTGTGCGCGGTGATACGGCTATTTCGTC 560  
DB 44088 ATTGTCTTATGCTGTGACCTTCGACAGGCGCCAGCTTTGTGTGACACGGCTGCTCTTC 44147  
QY 561 CTCGCTGTGCGCGCTTCACTGAGCTTGTGACAGCTTGTGCTCCGGGAAATCTCCACGCG 620  
DB 44148 CTCGCTGTGCGCGCTTCACTGAGCTTGTGACAGCTTGTGCTCCGGGAAATCTCCACGCG 44207  
QY 621 CCGGCTGTGCGCGCTTCACTGAGCTTGTGACAGCTTGTGCTCCGGGAAATCTCCACGCG 680  
DB 44208 GGTGCTGTGCGCGCTTCACTGAGCTTGTGACAGCTTGTGCTCCGGGAAATCTCCACGCG 44267  
QY 681 CCGGCGCTGTGCGCGGACGCTGCTGCAAGGCAATTTTCGCGGAGCGGATGAGGTTGCG 740  
DB 44268 GCACTATATGCGCGCGGACGCGCGCTGCAAGGCAATTTTCGCGGAGCGGATGAGGTTGCG 44327  
QY 741 ACGAGCGAAGGCTGCGCGCTGCTGCTGCAAGGCAATTTTCGCGGAGCGGATGAGGTTGCG 800  
DB 44328 CATGGGTGAGGCTGCTGCGCGCTGCTGCTGCAAGGCAATTTTCGCGGAGCGGATGAGGTTGCG 44387  
QY 801 CGATTCGATATTGCGCGCTGATTCAGAGATCCGCGATCAATCAAGCGGTGCGAGCAGCG 860  
DB 44388 TGAATCGATCTGCGCGGATCAAGGCTTCCGCAATCAATCAAGCGGTGCGAGCAGCG 44447  
QY 861 TCTGACGCTGCGAAGCGGACGCTCCCAAGAAATGATGCTGAAACGCGGCTTGGCGAGCG 920  
DB 44448 GCTGACACGCGCGGATGCGGCTTCCAGACAGGTTGTTGCTGACAGCGCTGACGCGCGC 44507  
QY 921 AGGCTGCGCGCGCTTCTGCTGAGGTTATGTCAGGACACACGCGACAGCGCTTGG 980  
DB 44508 TGGCATCTGCGCGCTGCGAGGATGAGTACGTGAAGCCACGCTTACGCGGACACGCGCTCG 44567  
QY 981 TGACCCCATCGAATTCAGCTTCAAGCTTCAATGCGCTTCCGCGGATGATGTCGAC 1040  
DB 44568 CGATTCGATGAGGTGAGGCGCTTCCGAGAGTATCCGCGGACGACGACGCTTCCGCGGC 44627  
QY 1041 GCGCGTGTGATCGGAGTCTGGAAGACCAACCTTGCGCATCTGAGTATGCGGTGAGT 1100  
DB 44628 TCCCTGTGCTGAGGTTCTGGAAGCCCAATCTGAGATACCGCGCGCGCGCGGCT 44687  
QY 1101 CACTGCGTGTGGAAGTCTGCTTGTCTCTTCAAGACGCGGAGATTCCTGCGCATCTTCA 1160  
DB 44688 CCGCGGCTGATCAAGTCTGCTCTCATCAAGAGGCGATTCCTCTCAACTGCA 44747  
QY 1161 CCGCGAGCGCTGAACCCCGGATCTCATGAGGATTCCTGCGCATCTTCA 1220  
DB 44748 CTACCGGAGCTCAACTCAATCTCTTGAAGAACCTCTGCTGATTCGAGAG 44807  
QY 1221 CCGAGACCGGTGCGGACGCTGATGACCGCGGACGCGGAGGCTGCTGCGCAT 1280  
DB 44808 GCTGATCTCTGCGCTGCGGAGAGGCGCGCTGATGACCGGATCACTGCTTCTGAGT 44867  
QY 1281 GAGCGGACCAACCGGACGCTGATGATGATGAGGCGCGGAGGCTGATGACCAACGCGC 1340  
DB 44868 GTACAGGACCAACCGGACGCTGATGATGAGGCGCGGAGGCTGATGACCAACGCGC 44927  
QY 1341 GCGCGCGGACGACCGGACGAGCTGCTGCTGCTGCGGAGGCTGATGACGCTTGA 1400  
DB 44928 GCGCGTGAAGGCTGCGGACGCTTCTGATGATGATGATGATGATGATGATGATGATGAT 44987  
QY 1401 TGCAAGGCGCGGCGCTGCGGACGATCTGATGATGATGATGATGATGATGATGATGAT 1460  
DB 44988 CGAGTTGCGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 45047  
QY 1461 TGTGCGTTGATGCTGCGAGACGCGCGACGCGATGAGACCGCGCTTCCGCGTGGCGC 1520

Db	45048	TCTCTGCTACTCGGGCGAAACCGGGGTGTGGAAGTTGAGCACTCTGGCGGTCCCCG	45107
Oy	1521	GACGTCGAGGAGGGGCTCGGGCGACGCTTGACGCTGCGGCGCAGAGGACAGCGTCGC	1580
Db	45108	GGGAGCACAAGCCAGATTGAGCGAGCGTCTCTCCGCGTTTCTTGAGGGGACAGCCGAAAC	45167
Oy	1581	CGGTGCGGTGGCGATATGCGCCGATTTCTCAACCGCGAAG---CTGGCTTTTCTTTCAC	1637
Db	45168	AGGCACTGTGACGGAAGGTCAAGGAGCGCCGTGTGGCCGGAGTGGCTTCTGTTCAC	45227
Oy	1638	CGGACAGGGGGCGGACGCTGGGCGATGGGCGCTGGGCTTACATGATATGATCCCGTT	1697
Db	45228	GGGGCAGGGGCTCGGAGTACCGGGGATGGGGCGGCACTTACCAAGACAGCCCAATT	45287
Oy	1698	CCGGGAGCGCTTCGACCTGTGCGTGGAGGCTTTCAACAGAGACTCGACCGGCGCTCCG	1757
Db	45288	CCGGCGTGCCTCGATCGGTGCGACGATGTGTGCGGCCCTTGTCTCGGGGCTCCATTCT	45347
Oy	1758	CGAGGTGATGTGGGCGGACCGGACAGCTCGACGCGGCTCTCGACCAAGACCTT	1817
Db	45348	GTCTGTATCTACCCCGAAGG-----GGGCAAGCTTCTCCATGACAGAGCGGCTT	45401
Oy	1818	CACCCAGCGGCGCTGTTCACCTTCGAATATGCGCTCGCGGCTGTGGGCGGTGAGGG	1877
Db	45402	CACGCAAGCCGCGTTGTTCGCGTTTCAAGTGGGCGCTCGGAACTCTGGGCTCTGTGGG	45461
Oy	1878	TGTAGAGCGGAGTTGTGTGCGCGGCGCAATAGCATTCGTAGACTGTGACTGCTGCGTGC	1937
Db	45462	CGTTGTACCGACCGTGTGATGGGCCCAAGCGTGGGTGATATGTGGCGGCGTGGCTGC	45521
Oy	1938	GGGCGGTGTTCTCGCTTGAAGACGCGGTGTTCTGTGTGCTGCGCGCGGCGCTGATGCA	1997
Db	45522	TGGGTGTTCACCTTCAGAAAGGAGGCTTGGCGTATCTGTGGAAGCGGCGCGGTGTATGCA	45581
Oy	1998	GGCGCTGCGGCGCGGGGGGAGATGTGTGATCGAGGCGCGGAGGCGGATGTGCTGC	2057
Db	45582	GTCGTTGCGCGGCGGCGGCGAGAGTGGCGGCTTTCACAGGAGGAGCTGTACCGCGCA	45641
Oy	2058	TGCGGTGGCGCGGACGACGCGTTCGATGCGATCGCGCGGTCAAGCTTCGGAACAGGT	2117
Db	45642	GGCATTGTGCTCTTAACGATCGCAGGTCTCATTTGCCGCTTCAGATGTCTTTCCGAGAC	45701
Oy	2118	GGTCACTGCGGGGCGCGGGCAACCTGTGCATGTGCATGCGGCGGCGCATGTGCGCGCGG	2177
Db	45702	GGTATCTCTTGGCATGGCAAAAGCGTGGAGGCGATTTCTGGGGCGTTTGTCCCGCAGAG	45761
Oy	2178	GGCGCGAACCAAGCGGCTCACGTCGTGCGATGCGTTCATCTCACGCTCATGTGCGCCGAT	2237
Db	45762	CGTBAAGTTTCGGGCGCTGGGTGTCTCCACGCGCTTTCACATCTTCGCTCATGTGACCCAT	45821
Oy	2238	GCTGAGGCGTTTCGGGCGTGTGTGCGCGAGTGTGAGCTACCGGCGGCGCTGCATCTGCT	2297
Db	45822	GCTGATGTGTTTCAAGCGGGGCGGGGAGAGACGCGTTATATCCCGCAAGATACGCT	45881
Oy	2298	GGTCACGATCTGACGAGGAGGCTTTCACAGACGAGGTGAGCTCTCGCGGCTATTGGT	2357
Db	45882	GATCTCGAACCTCACCGG-----TGGGCGCTCGGAGAGATTTCTCGGCGCGGCGTTTGG	45938
Oy	2358	GGCGCACGCGGAGAGGTGTGCGCTTCGCGATGTAGTGAAGGCGCTGCACGCGGCGG	2417
Db	45936	TCGACATGTGCGGTAAACCGGTCCGTTCTCGAATGGGATGGCGGCTCAAGGCAGTGG	45998
Oy	2418	TGCGGAGCATTGTCGAGGTGCGGTTCGGAATGCAAGCTCTCGGCGTGGTGTCTGCTG	2477
Db	45996	CGTCTCGGTCCGGGTGAGATTGGCCCGAATGCACTCTCTCGGATTCGGAAGCCGGTG	46055
Oy	2478	CATGCGGAGCGCCCGCGCGCGCTCTCGCATGTTCGCGCGCTGTGGCGTGAACGAGCGGCG	2537
Db	46056	TCTGCGGGAAGGTTCGGGTGTGTTGCTTCCCTTCGAAAGAGCAAAAGACGAGTGGGA	46115
Oy	2538	GACCGTGTGAGGCGCTCGGCGGCGTCTGTGGCCGTGTGCGCTGTCTCTGGGCGG	2597

D	b		46116	AGTCCTGCTGCACACGTGGTGCCTCTCTGTGCGAAGCAGCTCCGATCCGATCACTGGGACAAG	46117
O	y		2598	----CCTTTTCCTCTCAAGGGGGGGGGGGGGTCCGCTGCCACAGTAACCTTGACAGCGGA	2654
D	b		46176	GTTTCGACGAGCACTAACCCCGGTGGGGGGCTTCCTCCGCAATTATCTCTTTCAGACGA	46235
O	y		2655	GCCTACTCATGTAGTGCACGAAACCACGACGCGCGCGCTGGCGACCCGCGTCTCCGG	2714
D	b		46236	GCCTTTGCGCTC-----AACCAATGATGACGCGGGTGTGGCGCCGACGGTCAAG	46288
O	y		2715	AGCGGGTTCACAGCAGGTTCAGAGAGGGGGGCGCGGTGCGCGCGCGACCGCGCACGCG	2774
D	b		46289	TCCGGGTGCATCCCTTTGGATTCGCGCTGCGCTGCACGCGGTGACGCGCATCTGT	46348
O	y		2775	TCCGCTGACCATTCG 2790	
D	b		46349	TCCGATCGAAGTTCG 46364	
 RESULT 15 AR266876 . 3978 bp DNA linear PART 10-APR-2003					
LOCUS AR266876 Sequence 19 from patent US 6495348.					
ACCESSION AR266876					
VERSION AR266876.1 GI:29696258					
KEYWORDS Unknown.					
SOURCE Unknown.					
ORGANISM Unclassified.					
REFERENCE 1 (bases 1 to 3978)					
AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.					
TITLE Mitomycin biosynthetic gene cluster					
JOURNAL Patent : US 6495348-A 19 17-DEC-2002;					
FEATURES Location/Qualifiers					
source 1..3978					
BASE COUNT 497 a 1583 c 1415 g 483 t					
ORIGIN					
Query Match 19.1%; Score 814.2; DB 6; Length 3978;					
Best Local Similarity 57.6%; Pred. No.1,2e-79;					
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;					
O	y		18	CGAGCGCGCGACCGGAAGATCCGATTCGATCTCTCGACGAGTAGTTCGCTCCCGCGTGG	77
D	b		90	CGAGGAGCGCGCCCGCGACGCCGTCCGATCTGTGGCATGGCGTCCGCTTCCCGGGGA	149
O	y		78	CCTGATCGATCTGAGCGGGTTCGTGACGCTCTCGAGGGCTCGCGGACCACTCGGAGG	137
D	b		150	CGTGGCATTCGCGGACGACCTGTGGCAGCTGTGTCGCCGAGGGCCGGGAGCGCCTCACGA	209
O	y		138	AGTCCCGCGCGAAC--GCTGGAGTCAAGCAGCGGTGTTGATCCCGACCCCGATGCC	194
D	b		210	GTTCCTCCGCGGACCGGGGGTGGAGACGTGCAGCGCGCTAGACCCGACCGGGACCCC	269
O	y		195	GGGAGACAGCCCGGTACGCGCGCATCTTCTCTGACGACGTAGCTCTTGTGAAGCGCTC	254
D	b		270	GGGACGAGACCTAGCGCGCGCACCGCGGCTTCTCAAGAGACGCGCGCGGATTTGACGCCCC	329
O	y		255	CTTTTGGGATCTCGCTCGGAGACGCGTGGGAGGACCGTGCACATCGACTCTTGCT	314
D	b		330	CTTTTGGGATCAACGCGCGGAGGGGTCTCGCATGACCGGACAGCGCATGATCAT	389
O	y		315	GGAGGTGCTGGAGGCGCTGAGAGAAGCGCGCATCGCTCATCGGCGCTGCGGTAC	374
D	b		390	GGAGGTCTCTGGGAGGCGTTTCAAGCAGGCGGGCGCTCAAGCGACCACTCGGGGCGA	449
O	y		375	GGAAACGGAGATGTTCACTCGGATCGGCGCGCTCGAATAAGAGCCGCGCTCGGCACAC	434
D	b		450	GGAAGTGGGGGTCTTGCGTGGCTCCAACAGCAACGACTACATGATCAAGTGTCTCGACGC	509
O	y		435	GACGGCGTCCGACAGATGACGCTATGGCGGGGCTGGGAGCAATGCCAGGCTCGGAGC	494



Dh 510 GCGGAGCGTCGCGAGGCGTTT-----ATCGGAGACCGGCAACTCCGCGACGATCTCTC 563  
Qy 495 GGGCGGATCTCGATGCGCTCGGGAGTCGAGGGCCGCTGCTCGGGGATACGAGCTTA 554  
Dh 564 CGGCGCGGCTGCGCTTACACCTTCGAGCTTCGAGGCGCGGCGCTGCTCGAGCACGCGCTG 623  
Qy 555 TTGCTCTCGTGTGCGGCTTCACTTGGCTGTCAAGCTTGTGCTTCGCGGAAATGTC 614  
Dh 624 CTCTCTCTCGTGTGCGGCTGACCTTGGCGGAGCTCTGCGGAGGGGAGTGCTC 683  
Qy 615 CAGCGGCTGTGCTGTGAGGATGCTGTATGCTGTGCGGAGACCTCTGTGTGCTTC 674  
Dh 684 CTGTGCTGTGCGGCGGCGGAGCGGTATGCGACCGCTTCACTGAGTTCAAG 743  
Qy 675 GAAAGCCCGGCGGCTGTGCGGAGGAGCGGTGCTGACGAGCATTTTTCGCGGAGGCGGATG 734  
Dh 744 CCGGCAAGCGGCGCTGTGCGGCGGCGGCGGCTGTGAGTCTTCTGTGCGAGCGCGCGAG 803  
Qy 735 GTTTCGAGGAGGCGGAGGCGGCGGCTGTGCTGTGCTGTGAGGCGCTCACTGTGAGCGCGCG 794  
Dh 804 CACCACTGTGTCCGAGGCGGCGGCGGCTGTGCTGTGCGGCGGCTGTGAGCGCGCGCG 863  
Qy 795 GGAAGCGGATGCGATATTGCGGCTGTGATTCGAGGATCCGCGATCAATCAAGCGGTGCGAG 854  
Dh 864 CTTGGGCTTACCCCGTGTGACGCGGTATCTCGGAGGAGCGCGTCAACAGAGCGGCGGAG 923  
Qy 855 CAGCGGTCTGACCGGCGGAGAGGAGCTCCCAAGAAATGCTGTGCGAAAGGCGGCGCTGGC 914  
Dh 924 CCGGCGCTGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 983  
Qy 915 GGAAGCGGCTGTGCGGCGGCGGCGGCTGTGCTGTGCTGTGAGGCGGCGGCGGCGGCGG 974  
Dh 984 CAAAGCAAGCTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1043  
Qy 975 GCTTGTGTGATCCCATTCGAATTCAGCTGTGAATGCGGTATCGGCTCGGCGGAGATGT 1034  
Dh 1044 GCTGTGGGAGCCCGATTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1103  
Qy 1035 CGGCGGCGGCTGTGATCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1094  
Dh 1104 GGGGAGGCGGCTGTGTGCTGTGCTGTGATGATTCGAATTCGAGGCGGCGGCGGCGG 1163  
Qy 1095 GGGGATCACTGAGGCTGTGAGGCTGTGCTGTGCTGTGCTGTGAGGCGGCGGCGGCGG 1154  
Dh 1164 CGGCGGCGGCGGCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1223  
Qy 1155 CTTCAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1214  
Dh 1224 GCTGCACTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1283  
Qy 1215 GCGGCGGCGGAGCACCGTGTGCGGAGCTGGAATACCGCGGAGCGGCGGCGGCGGCGGCT 1274  
Dh 1284 GACGAGGCGGAGGAGCTGTGCGGAGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCT 1343  
Qy 1275 CGGAGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1334  
Dh 1344 CGGAGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1403  
Qy 1335 ACCGCGG-----GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1376  
Dh 1404 GGAAGCGGCGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1463  
Qy 1377 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1436  
Dh 1464 CGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1523  
Qy 1437 CTACCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1496  
Dh 1524 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1583  
Qy 1497 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1556  
Dh 1584 CGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1643

Qy 1557 TGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1616  
Dh 1644 CTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1703  
Qy 1617 CAAAGCTGCTTCTTCTTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1676  
Dh 1704 CAAAGGCGGCTTCTTCTTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1763  
Qy 1677 GTAGATGTATGTGCGGCTTCTTCTTCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1726  
Dh 1764 GCGGCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1823  
Qy 1737 GGAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1796  
Dh 1824 GCAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1883  
Qy 1797 GCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1856  
Dh 1884 CTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1943  
Qy 1857 CGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1916  
Dh 1944 GCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2003  
Qy 1917 GCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1976  
Dh 2004 ACTGAGCGGCGGCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2063  
Qy 1977 TGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2036  
Dh 2064 CGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2123  
Qy 2037 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2096  
Dh 2124 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2183  
Qy 2097 GGTCAAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2156  
Dh 2184 GGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2243  
Qy 2157 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2216  
Dh 2244 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2303  
Qy 2217 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2276  
Dh 2304 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2363  
Qy 2277 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2336  
Dh 2364 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2423  
Qy 2337 GAGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2396  
Dh 2424 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2483  
Qy 2397 GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2456  
Dh 2484 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2543  
Qy 2457 GCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2513  
Dh 2544 GAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2603  
Qy 2514 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2573  
Dh 2604 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2663  
Qy 2574 CGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2633  
Dh 2664 CGGCTTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2723



Qy 2634 CACGTACCTTGGACAGCGGCTACTGATCGACAGAAAGCCG 2680  
Db 2724 CACGTACGCTTTCACAGCGGCCACTACTGCTGCGGCGATGACCg 2770

Search completed: October 4, 2003, 03:30:21  
Job time : 15291.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 125.882 Seconds  
(without alignments)  
2912.986 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210  
Sequence: 1 VADPRIERADPIAIVGAS.....GVNDPVSSGADQDWETIAL 1421

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7207	100.0	1421	2	Q9L8C9
2	7081	98.2	1421	2	Q9KJ00
3	4137	57.4	2439	2	Q9L8C5
4	4098	56.8	2439	2	Q9K1Z5
5	3322.5	46.1	7257	2	Q9K1Z7
6	3311.5	45.9	7257	2	Q9L8C7
7	2634	36.5	2218	2	Q8BJY1
8	2606	36.1	6315	2	Q9ADL6
9	2416.5	33.5	5017	2	Q8GBX6
10	2357	32.7	5644	2	Q9JNX8
11	2355.5	32.7	5435	2	Q9JNX2
12	2331.5	32.3	11096	2	Q9L4W3
13	2330	32.3	9507	2	Q9EMW1
14	2328.5	32.3	10917	2	Q9JNW6
15	2321	32.2	4151	16	Q53450
16	2313	32.1	3513	2	Q8GBX5

17	2311.5	32.1	6146	2	Q9JHJ5	Q9JHJ5 streptomyc
18	2298.5	31.9	2103	16	Q9CD81	Q9CD81 mycobacteri
19	2275.5	31.6	1259	2	Q8RJX8	Q8RJX8 stigmatella
20	2273	31.5	2152	2	Q9ALM5	Q9ALM5 saccharopol
21	2263.5	31.4	1885	2	Q8RJY4	Q8RJY4 stigmatella
22	2260.5	31.4	9477	2	Q9L4X3	Q9L4X3 streptomyc
23	2260	31.3	8817	2	Q93840	Q93840 polyanthum
24	2258	31.3	9510	2	Q9JNX9	Q9JNX9 streptomyc
25	2257	31.3	1604	2	Q8RJX9	Q8RJX9 stigmatella
26	2255.5	31.3	1585	2	Q8RJY5	Q8RJY5 stigmatella
27	2249	31.2	6145	2	Q9JHJ4	Q9JHJ4 streptomyc
28	2245.5	31.1	3798	2	Q9L8C6	Q9L8C6 polyanthum
29	2238.5	31.0	3798	2	Q9K1Z6	Q9K1Z6 polyanthum
30	2229.5	30.9	5192	2	Q9JTW9	Q9JTW9 stigmatella
31	2211.5	30.7	2478	16	Q8JTN5	Q8JTN5 anabaena sp
32	2207	30.6	1602	16	Q65933	Q65933 mycobacteri
33	2199.5	30.5	3352	2	Q9JHJ3	Q9JHJ3 streptomyc
34	2174	30.2	3579	2	Q8GBX4	Q8GBX4 polyanthum
35	2165	30.0	6797	2	Q9X993	Q9X993 streptomyc
36	2157.5	29.9	1360	2	Q9RFK6	Q9RFK6 stigmatella
37	2157.5	29.9	3524	2	Q9JHJ6	Q9JHJ6 streptomyc
38	2156.5	29.9	6048	2	Q9JHJ7	Q9JHJ7 streptomyc
39	2152	29.8	1835	2	Q9JHJ5	Q9JHJ5 streptomyc
40	2150	29.8	1841	2	Q9JHJ5	Q9JHJ5 streptomyc
41	2136	29.6	2518	16	Q8JTN4	Q8JTN4 anabaena sp
42	2128	29.5	3816	2	Q9K1V3	Q9K1V3 streptomyc
43	2127	29.5	3170	2	Q9ALM4	Q9ALM4 saccharopol
44	2123.5	29.5	2024	2	Q9EMW3	Q9EMW3 streptomyc
45	2120	29.4	1587	16	Q8JYN5	Q8JYN5 anabaena sp

## ALIGNMENTS

RESULT 1  
Q9L8C9 PRELIMINARY; PRT; 1421 AA.  
ID Q9L8C9  
AC Q9L8C9  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Polylactide synthase.  
GN EPOA.  
OS Polyanthum cellulorum.  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
OC Sorangineae; Polyanthaceae; Polyanthum.  
OX NCBI\_taxid=56;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=So ce90;  
RX MEDLINE=20130945; PubMed=10662695;  
RA Molnar I., Schupp T., Ono M., Zirkle R.E., Milnamow M.,  
RA Nowak-Thompson B., Engel N., Toupet C., Stratzmann A., Cyr D.D.,  
RA Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;  
RT "The biosynthetic gene cluster for the microtubule-stabilizing agents  
epothilones A and B from Sorangium cellulorum So ce90.";  
RL Chem. Biol. 7:97-109 (2000).  
DR EMBL: AF210843; AAF26919.1; -  
DR HSRP; P25715; 1MLA.  
DR InterPro: IPR001227; AC transferase.  
DR InterPro: IPR002085; Adh\_zn family.  
DR InterPro: IPR000794; Ketoacyl-synt.  
DR InterPro: IPR006162; Pantone attach.  
DR InterPro: IPR006163; PP bind.  
DR Pfam: PF00698; Acyl\_transf. 1.  
DR Pfam: PF00107; adh\_zinc. 1.  
DR Pfam: PF00109; ketoacyl-synt. 1.  
DR Pfam: PF02801; ketoacyl-synt C; 1.  
DR Pfam: PF00550; pp-binding. 1.  
DR PROSITE: PS50075; ACP DOMAIN: 1.  
DR PROSITE: PS00012; PROSOPHANTHREINE. 1.  
DR Phosphatetheine.  
KW SEQUENCE 1421 AA; 149035 MW; ABB5615EDAEDC996 CRC64;

Query Match 100.0%; Score 7207; DB 2; Length 1421;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1420; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRPETRAEDPIALVAGASCRLPGVYIDLSGFMTLLSGSRDVTGVRPAPERMMAAMFDP 60  
 1 MDRPIETRAEDPIALVAGASCRLPGVYIDLSGFMTLLSGSRDVTGVRPAPERMMAAMFDP 60  
 DB 61 DDPAPCKTPTVTRASFISDVACPDASFPGISPREALMDPAHRLLEVCHEALENAAIAS 120  
 61 DDPAPCKTPTVTRASFISDVACPDASFPGISPREALMDPAHRLLEVCHEALENAAIAS 120  
 QY 121 ALVGTETGVFTIGSPSEYEALPOATASAEIDAHGGIGTWPAGRISVALGRPCVA 180  
 121 ALVGTETGVFTIGSPSEYEALPOATASAEIDAHGGIGTWPAGRISVALGRPCVA 180  
 DB 181 VDTAVSSSLVAHVLACQSLRSGECSTALAGVSLMLSPSTLWLSKTRALARDGRKAFS 240  
 181 VDTAVSSSLVAHVLACQSLRSGECSTALAGVSLMLSPSTLWLSKTRALARDGRKAFS 240  
 QY 241 AEAADGFRGEGCAVVLKRLSGARADGRIILAVIRGSAINHDCASSGLTVPNGSSQETVL 300  
 241 AEAADGFRGEGCAVVLKRLSGARADGRIILAVIRGSAINHDCASSGLTVPNGSSQETVL 300  
 DB 241 AEAADGFRGEGCAVVLKRLSGARADGRIILAVIRGSAINHDCASSGLTVPNGSSQETVL 300  
 301 KRALLADGCAASSVGVYEAHGTGTLGDPTEIQALNAVGLGRDVAITPILIGSVKTNLGH 360  
 301 KRALLADGCAASSVGVYEAHGTGTLGDPTEIQALNAVGLGRDVAITPILIGSVKTNLGH 360  
 QY 361 PEVASGITGLLKVLSLQHGQIPAHLLAQLNPRISWGDRLITVTRAPTPMDMTPRRA 420  
 361 PEVASGITGLLKVLSLQHGQIPAHLLAQLNPRISWGDRLITVTRAPTPMDMTPRRA 420  
 DB 361 PEVASGITGLLKVLSLQHGQIPAHLLAQLNPRISWGDRLITVTRAPTPMDMTPRRA 420  
 421 GVSSFGMSGTNAVLEAPAACTPPAPERPAPBLVLSARTKSALDAQARLDHLETY 480  
 421 GVSSFGMSGTNAVLEAPAACTPPAPERPAPBLVLSARTKSALDAQARLDHLETY 480  
 QY 481 PSCGLDGVAFSLTTSAMEHRLAVATSEGRALMDAAQOTSGAARSTIADSSRG 540  
 481 PSCGLDGVAFSLTTSAMEHRLAVATSEGRALMDAAQOTSGAARSTIADSSRG 540  
 DB 481 PSCGLDGVAFSLTTSAMEHRLAVATSEGRALMDAAQOTSGAARSTIADSSRG 540  
 541 LAFLFTGQAGQTLGMRGLYDVSAFREAFDLCVRLFNQELDRPLREVMMAEPASVDAL 600  
 541 LAFLFTGQAGQTLGMRGLYDVSAFREAFDLCVRLFNQELDRPLREVMMAEPASVDAL 600  
 QY 601 LDDTAFTQPALFTFEVYLAALMSWGVPELVAGHSGELVAAQVAGVSELEDAVFLVAA 660  
 601 LDDTAFTQPALFTFEVYLAALMSWGVPELVAGHSGELVAAQVAGVSELEDAVFLVAA 660  
 DB 601 LDDTAFTQPALFTFEVYLAALMSWGVPELVAGHSGELVAAQVAGVSELEDAVFLVAA 660  
 661 RGRIMQALPAGAMVSTIEAPEADVAALVAAPHAASVSTIAAVNAPQOVVITAGAGOPVHAIA 720  
 661 RGRIMQALPAGAMVSTIEAPEADVAALVAAPHAASVSTIAAVNAPQOVVITAGAGOPVHAIA 720  
 QY 721 AMARAGARTALVSHAHSPLMAPLEAFGVAESVYRPSIVLVSNLGKACTDEVS 780  
 721 AMARAGARTALVSHAHSPLMAPLEAFGVAESVYRPSIVLVSNLGKACTDEVS 780  
 DB 721 AMARAGARTALVSHAHSPLMAPLEAFGVAESVYRPSIVLVSNLGKACTDEVS 780  
 781 SPGYWRHAREVVRPADGVYALHAAGAGTVEVGPSTLGLVPCMPDARPALASSRA 840  
 781 SPGYWRHAREVVRPADGVYALHAAGAGTVEVGPSTLGLVPCMPDARPALASSRA 840  
 QY 841 GRDEPAPVLEALGLMVAVGLVSWAGLPSPGGRVPLPTYPWOKERWIDTKADDDARG 900  
 841 GRDEPAPVLEALGLMVAVGLVSWAGLPSPGGRVPLPTYPWOKERWIDTKADDDARG 900  
 DB 841 GRDEPAPVLEALGLMVAVGLVSWAGLPSPGGRVPLPTYPWOKERWIDTKADDDARG 900  
 901 RRAFGAGHDEVEEGAVRGDRSARLDHPSPSGREKTEAAGDRPFRLIEIDEPVLDH 960  
 901 RRAFGAGHDEVEEGAVRGDRSARLDHPSPSGREKTEAAGDRPFRLIEIDEPVLDH 960  
 QY 961 LVRVTRERRAPGIGVEIIVADAAGLSNDVOLALGMVDDLPKGNPPLLIGECACRIV 1020  
 961 LVRVTRERRAPGIGVEIIVADAAGLSNDVOLALGMVDDLPKGNPPLLIGECACRIV 1020  
 DB 961 LVRVTRERRAPGIGVEIIVADAAGLSNDVOLALGMVDDLPKGNPPLLIGECACRIV 1020

QY 1021 AVEGCVNGLVGQPIYALISAGAFATHVTTSAALVPRPOALSAIEAAMPVAYLTAWYL 1080  
 1021 AVEGCVNGLVGQPIYALISAGAFATHVTTSAALVPRPOALSAIEAAMPVAYLTAWYL 1080  
 DB 1081 DRIRLOPGERVLIIAATCGVGLAAVQMAQHVGAEVHATAGTPERKAYLESIGRVYS 1140  
 1081 DRIRLOPGERVLIIAATCGVGLAAVQMAQHVGAEVHATAGTPERKAYLESIGRVYS 1140  
 QY 1141 RSDRFADVRAWTGEGVDVVLNSLSGELIDSFNLSHGRFVELGKRDCAVDNQLGR 1200  
 1141 RSDRFADVRAWTGEGVDVVLNSLSGELIDSFNLSHGRFVELGKRDCAVDNQLGR 1200  
 DB 1141 RSDRFADVRAWTGEGVDVVLNSLSGELIDSFNLSHGRFVELGKRDCAVDNQLGR 1200  
 1201 PFLRNLSFSLVDRGWMLEPAPRVALLLEELGLIAAGVTPPPIATLPIAVADAFRSM 1260  
 1201 PFLRNLSFSLVDRGWMLEPAPRVALLLEELGLIAAGVTPPPIATLPIAVADAFRSM 1260  
 QY 1261 AQAHLGKLVLTGDPVOIRIPTAGACPSGTGRDLDRLASAAPARAALLLEAFLRQ 1320  
 1261 AQAHLGKLVLTGDPVOIRIPTAGACPSGTGRDLDRLASAAPARAALLLEAFLRQ 1320  
 DB 1261 AQAHLGKLVLTGDPVOIRIPTAGACPSGTGRDLDRLASAAPARAALLLEAFLRQ 1320  
 1321 VSOVARTPEIKVGAELFRLGMDSLMAVELRNRIEASIKLSTFTSTSPNIALAON 1380  
 1321 VSOVARTPEIKVGAELFRLGMDSLMAVELRNRIEASIKLSTFTSTSPNIALAON 1380  
 QY 1381 LLDALATALLERVAENIRAGVONDFVSSGADQDWEIATL 1421  
 1381 LLDALATALLERVAENIRAGVONDFVSSGADQDWEIATL 1421  
 DB 1381 LLDALATALLERVAENIRAGVONDFVSSGADQDWEIATL 1421

RESULT 2  
 Q9KJ00 PRELIMINARY; PRT; 1421 AA.  
 ID Q9KJ00;  
 AC Q9KJ00;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE EgoA.  
 GN EPOA.  
 OS Polyangium cellulosum.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 OC Sorangineae; Polyangiaceae; Polyangium.  
 OX NCBI\_TaxID=56;  
 RN [1] -  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SMP44;  
 RX MEDLINE=20293058; PubMed=10831849;  
 RA Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;  
 RT "Isolation and characterization of the epoIII gene cluster from Sorangium cellulosum.";  
 RL Gene 249:153-160(2000).  
 RU [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SMP44;  
 RX MEDLINE=20115953; PubMed=10649995;  
 RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C.;  
 RT "Cloning and heterologous expression of the epoIII gene cluster.";  
 RL Science 287:640-642(2000).  
 RU EMBL; AF217189; AAF62880.1; -.  
 DR HSP; P25715; IMLA.  
 DR InterPro: IPR001227; AC transferase.  
 DR InterPro: IPR002085; Adh\_zn\_family.  
 DR InterPro: IPR000794; Ketoacyl-synt.  
 DR InterPro: IPR006162; Pantane attach.  
 DR InterPro: IPR006163; PP\_bind.  
 DR Pfam: PF00698; Acyl\_transf.1.  
 DR Pfam: PF00107; adh\_zinc.1.  
 DR Pfam: PF00109; ketoacyl-synt.1.  
 DR Pfam: PF02801; ketoacyl-synt.C.1.  
 DR Pfam: PF00550; pp-binding.1.  
 DR PROSITE: PS50075; ACP\_DOMAIN.1.  
 DR PROSITE: PS00012; PHOSPHOPANTHETINE.1.  
 DR Phosphoanettheine.  
 SQ SEQUENCE 1421 AA; 148744 MW; 3D055DDBF686E2C5 CRC64;

Query Match 98.2%; Score 7081; DB 2; Length 1421;  
 Best Local Similarity 98.2%; Pred. No. 0;  
 Matches 1396; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

1 VADRPTEAAEDPIAIVGASCRIPGVITLSCGFTLLGSGRTDVGVPAREMDAAEDP 60  
 1 MADRPTEAAEDPIAIVGASCRIPGVITLSCGFTLLGSGRTDVGVPAREMDAAEDP 60  
 61 DDAPKPTVTNATSPISDVACPDSPFISPREALMDPAHRLLEVCMALENAAIAS 120  
 61 DDAPKPTVTNATSPISDVACPDSPFISPREALMDPAHRLLEVCMALENAAIAS 120  
 121 ALVGETGVFIGIGPSEYEALPQATASAEIDAHGGLGTMPSVAGARISYALGRPCVA 180  
 121 ALVGETGVFIGIGPSEYEALPQATASAEIDAHGGLGTMPSVAGARISYALGRPCVA 180  
 181 VDTAVSSSLVAHVALCQSLRSGECSTALAGVSLMSPSTLWLSKTRALATDGRKAPS 240  
 181 VDTAVSSSLVAHVALCQSLRSGECSTALAGVSLMSPSTLWLSKTRALATDGRKAPS 240  
 241 AADGGRGEGCAVVVVKRLSGARADGRIIAYIRGSAINHOGASSGLTVPNGSSQEV 300  
 241 AADGGRGEGCAVVVVKRLSGARADGRIIAYIRGSAINHOGASSGLTVPNGSSQEV 300  
 301 KEALADAGCAASSVGVYEAHGTGTTGDPRIEIOALNAVYGLGRDVATPLLIGSVKTNLGH 360  
 301 KEALADAGCAASSVGVYEAHGTGTTGDPRIEIOALNAVYGLGRDVATPLLIGSVKTNLGH 360  
 361 PEYASGITGLLKVVLSTLQHGQIPAHHAQALNPRISMGDLRLTVTRARTPMDMTPRRA 420  
 361 PEYASGITGLLKVVLSTLQHGQIPAHHAQALNPRISMGDLRLTVTRARTPMDMTPRRA 420  
 421 GVSSFGMSGTNAHVLEEAIPAATCTPAPERPPELLVLSARTASALDAQAARLDLETTY 480  
 421 GVSSFGMSGTNAHVLEEAIPAATCTPAPERPPELLVLSARTASALDAQAARLDLETTY 480  
 481 PSQCLGDVAFSLATTSAMEHRLAVALATREGLRALDAQAQOTSPGAVRSIADSSRQK 540  
 481 PSQCLGDVAFSLATTSAMEHRLAVALATREGLRALDAQAQOTSPGAVRSIADSSRQK 540  
 541 LAFLEFGGAGQTLGMGRGLYDVWASAFREAFDLCVRLFNQELDRPLEVMAEBSVDAAL 600  
 541 LAFLEFGGAGQTLGMGRGLYDVWASAFREAFDLCVRLFNQELDRPLEVMAEBSVDAAL 600  
 601 LDQTAFTQALPFFFEVALALMRSWGPBEELVAGHSIGELVACVAGVSLDEAVFLVA 660  
 601 LDQTAFTQALPFFFEVALALMRSWGPBEELVAGHSIGELVACVAGVSLDEAVFLVA 660  
 661 RGRLMQALPAGGAMVSIIEAPADVAAVAAPHAASVSIIAVNGPDQVVIAGAGOPVAAIA 720  
 661 RGRLMQALPAGGAMVSIIEAPADVAAVAAPHAASVSIIAVNGPDQVVIAGAGOPVAAIA 720  
 721 AMAARGARTKALHVSHPMLAEMLEAFGRVAESVYRPSIVLVSMLSGKACTDEVS 780  
 721 AMAARGARTKALHVSHPMLAEMLEAFGRVAESVYRPSIVLVSMLSGKACTDEVS 780  
 781 SGQWVRHAREVVRFPADGVKALHAAGAFVEVGPSSTLLGLVPACMPDARPLLASSSA 840  
 781 SGQWVRHAREVVRFPADGVKALHAAGAFVEVGPSSTLLGLVPACMPDARPLLASSSA 840  
 841 GRDEPATVLEALGGLVAVGLVSMAGLFRSGGRVPLPTYPMDREKRYVITDTRADDAARD 900  
 841 GRDEPATVLEALGGLVAVGLVSMAGLFRSGGRVPLPTYPMDREKRYVITDTRADDAARD 900  
 901 RRAFGAGHDEVEEGAVRGDRRSARLDHPPEPSGRREKVEAAGDBPFLIEIDEPGLDH 960  
 901 RRAFGAGHDEVEEGAVRGDRRSARLDHPPEPSGRREKVEAAGDBPFLIEIDEPGLDH 960  
 961 LVLRVTERAPGIGVEIIVADAAGLSFNDVQALAGVPPDLPGKPNPPLLIGECGCRIV 1020  
 961 LVLRVTERAPGIGVEIIVADAAGLSFNDVQALAGVPPDLPGKPNPPLLIGECGCRIV 1020

1021 AVEGNGVGVQPIYALASAGAFATHVTSALVLRPQALSAIEAAMPVAVLTAWVAL 1080  
 1021 AVEGNGVGVQPIYALASAGAFATHVTSALVLRPQALSAIEAAMPVAVLTAWVAL 1080  
 1081 DRIALQGERLILHAATGCGVLAIVQMAOHVGAHVATAGTEPKRAYLESIGVRYVSS 1140  
 1081 DRIALQGERLILHAATGCGVLAIVQMAOHVGAHVATAGTEPKRAYLESIGVRYVSS 1140  
 1141 RSDRFVADVRAWTGEGVDVNVNSLSEGLIDSFWLRSRGRFVEIGKDCYADNOLGR 1200  
 1141 RSDRFVADVRAWTGEGVDVNVNSLSEGLIDSFWLRSRGRFVEIGKDCYADNOLGR 1200  
 1201 PFLRNLSFSLVDRGMLEERPARVALLEBLGLIAAGVTPPPIATLPIAVADAFRSM 1260  
 1201 PFLRNLSFSLVDRGMLEERPARVALLEBLGLIAAGVTPPPIATLPIAVADAFRSM 1260  
 1261 AQAQHLGKLVTLGPEVQIRPTAGAGPSGDRDLDRLASAAPAAALAEFLRQ 1320  
 1261 AQAQHLGKLVTLGPEVQIRPTAGAGPSGDRDLDRLASAAPAAALAEFLRQ 1320  
 1321 VSQVLRTEPKVGAELFTRLGMDSLMAVELNRLEASIKLSTTFSTSPNIALTON 1380  
 1321 VSQVLRTEPKVGAELFTRLGMDSLMAVELNRLEASIKLSTTFSTSPNIALTON 1380  
 1381 LLDALATLSTLERRVAENLRAGQNDPVSSGADQDMEIATL 1421  
 1381 LLDALATLSTLERRVAENLRAGQNDPVSSGADQDMEIATL 1421

RESULT 3  
 ID 0918CS PRELIMINARY; PRT: 2439 AA.

AC 0918CS: 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Polyketide synthase.  
 GN EPOE.  
 OS Polyangium cellulosum.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 OC Sorangium; Polyangiaceae; Polyangium.  
 OX NCBI\_Taxid=56;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=So ce90;  
 RX MEDLINE=20130945; PubMed=10662695;  
 RA Molnar I., Schnupp T., Ono M., Zirkle R.E., Milamow M.,  
 RA Nowak-Thompson B., Engel N., Toupet C., Straumann A., Cyr D.D.,  
 RA Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;  
 RT "The biosynthetic gene cluster for the microtubule-stabilizing agents  
 RT epothilones A and B from Sorangium cellulosum So ce90.";  
 RL Chem. Biol. 7:97-109(2000).  
 DR EMBL: AF210843; AAF26923.1; -.  
 DR HSSP: P251745; IMLA.  
 DR InterPro: IPR001227; AC transferase.  
 DR InterPro: IPR002085; Adh. zn. family.  
 DR InterPro: IPR00794; Ketocacyl-synt.  
 DR InterPro: IPR006162; Pentane\_attach.  
 DR InterPro: IPR001031; Thioesterase.  
 DR pfam: PF00698; Acyl\_transf. 1.  
 DR pfam: PF00107; adh\_zinc. 1.  
 DR pfam: PF00109; ketocacyl-synt. 1.  
 DR pfam: PF02801; ketocacyl-synt\_C. 1.  
 DR pfam: PF00550; pp-binding. 1.  
 DR pfam: PF00975; Thioesterase. 1.  
 DR PROSITE: PS50075; ACP\_DOMAIN. 1.  
 DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE. 1.  
 DR PROSITE: PS00012; PHOSPHOPANTHERINE. 1.  
 KM Phosphopantetheine; Transferase.  
 SQ SEQUENCE 2439 AA; 257732 MW; 08A32D75BCBADF2C CRC64;

Query Match

57.4%; Score 4137; DB 2; Length 2439;



```

RESULT 4
O9K125 PRELIMINARY; PRT; 2439 AA.
AC O9K125;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE EPOF.
GN EPOF.
OS Polyanthum cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangineae; Polyangliaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SMp44;
RX MEDLINE=20293058; PubMed=10831849;
RA Julien B., Shah S., Ziemann R., Goldman R., Katz L., Khosla C.;
RT "Isolation and characterization of the epoIII gene cluster from Sorangium cellulosum.";
RL Science 249:153-160(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SMp44;
RX MEDLINE=20115953; PubMed=10649995;
RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT "Cloning and heterologous expression of the epoIII gene cluster.";
RL Science 287:640-642(2000).
DR EMBL; AF217189; AAF62885.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR000794; Ketocyl-synt.
DR InterPro; IPR006162; Ppantn_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00698; Acyl_cranf; 1.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketocyl-synt; 1.
DR Pfam; PF02801; ketocyl-synt_C; 1.
DR Pfam; PF00550; pp-binding; 1.
DR Pfam; PF00975; Thioesterase; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 1.
DR KEGG; Phosphopantetheine; Transferase.
SQ SEQUENCE 2439 AA; 257507 MW; D3A972BC49FC3497 CRC64;

Query Match 56.8%; Score 4098; DB 2; Length 2439;
Best Local Similarity 42.8%; Pred. No. 3.2e-192;
Matches 922; Conservative 139; Mismatches 315; Indels 776; Gaps 11;

```

```

OY 305 ADAGCAASSGVYEAHGTGTTGDDPIEIOALNAVYGLGADVATPLLIGSVKTNLGHPEYA 364
DB 318 QSAVADAGAIIDVEYHGTGTSIGDDPIEVDALNAVGVPAADSRCLGAVKTNLGLHEBA 377
OY 365 SGITGLKVLVSLQHQIIPAILHAQALNPRISWGLRLTVTRAPTPWPMNTPRRAVSS 424
DB 378 AGVAGIKKAAALHHEISIPRNHFTPLNPRILEGTALALATEPVPWPAQPRFAGVA 437
OY 425 PGMGTNNAVLEEA PAATCTPPAPERPELLVLSARTSALDAQARLDLETYPSC 484
DB 438 FGLSGTNNAVLEEA PATVLA PATGSRSELVLVLSAKSTAALDAQARLSAHIAAPEG 497
OY 485 LGDVAFSLATTSAAEHRLAVALATSGREGIALDLAAQOGTSPGAVRSIADSSRGTAFL 544
DB 498 LGDVAFSLVATTSAPMEHRLAVALATSGREGIALDLAAQOGTSPGAVRSIADSSRGTAFL 557
OY 545 FTGGAGQTLGMRGLYDVWSAFREAPDLCTVRLFNQELDPLREVMKAEPAVDALLDOT 604
DB 558 FAGGAGQVVGMRGLMEAMPARETFEDRCVTLFDEHLQPLCEVWMAEFGSSRSLLDOT 617
OY 605 AFTOPALTFPEVALAALMRSGVEPELVAGHSIGELVAACVAGVSLIEDAVFLVARGL 664
DB 618 AFTOPALTFEVALAALFRSGVEPELVAGHSIGELVAACVAGVSLIEDAVFLVARGL 677
OY 665 MOALPAGAMVSI EAPEADVAALVA PHAASVIAAVNAPDQVVIAGAGQPVHIAAAMMA 724
DB 678 MOALPAGAMVSI EAPEADVAALVA PHAASVIAAVNAPDQVVIAGAGQPVHIAAAMMA 737
OY 725 RGARTKALVSHA FSPPLMA PML EAFGRVAESVSYRPSIVLVSNLSGACTDEVS PGY 784
DB 738 RGARTKPLVSHA FSPPLMD PML EAFRRVTESVYRPSMALVSNLSGACTDEVA PGY 797
OY 785 WVRHAREVVRADGVKALHAAGCTFVEGPESTLLGVPACMPARPALASSRGRBE 844
DB 798 WVRHAREVVRADGVKALHAAGCTFVEGPESTLLGVPACMPARPALASSRGRBE 857
OY 845 PATVLEALGGLWAGLVSMAGLPSSGGRVPLPTYPMQREKWI DTKADDAARGRRA- 903
DB 858 AASALEALGFPVVGSGVTVMSGVFSGGRVPLPTYPMQREKWI EAPVDSAGDQIGRAG 917
OY 904 ----- PGACH----- 908
DB 918 AGDHPILGEAFSVTHAGRLMETTLDRKRLPWLGEHRAQGVVFPAGLYLMASSGAE 977
OY 909 ----- 908
DB 978 ILGDGPQVTVLLETLPFAGDTAVPVQVVTTEERPGRLPQVARSREBARAFRIHA 1037
OY 909 ----- 908
DB 1038 RGVLRVGRAEFPALNLAALRAHRAVPAALITGALAEMKLYGPRALGLAELMRGEG 1097
OY 909 ----- 908
DB 1098 BALGRVRLPESAGSATAYQLHVLVDACVQIVGA FADRBDEATPMAVVEGVRLEFORSP 1157
OY 909 ----- 908
DB 1158 GELMCHARVSDGQA PSRWSAD FELMDGTGA VAEISRLVVERLSAGVRRRADDWFLE 1217
OY 909 ----- 908
DB 1218 LDMEPALLEGPRITAGRMILLQEGGLGRSLCSALKAAGHVVVHAAGDTSAAAGRALLA 1277
OY 909 ----- 908
DB 1278 NAFDQAPTA VVHLSSLDGGQLDPLGAGALDAPRPVDADALLESALMRGCDVSL 1337
OY 909 ----- 908
DB 1338 VOALVGMILRNAPRLMLLTRGAQAAAAGDVSVVQAPLLGLGRTIALBHAEALRCISVDLP 1397

```

QY 909 -----DEVEEGAVRGDRRSARLDHPPESSGREKVEAAGDPFPLEI 952  
| : : : : :  
Db 1398 ACPGEADALLAELADDAEEVALRGGERFVALVRLLEAQRREKIAAGRPFFLEI 1457  
QY 953 DEFGVDHLVLTERRAPGLGEVEIANDAGLSFNDVOLALGMVPDDLPKPNPILLG 1012  
| : : : : :  
Db 1458 DEFGVDLQVLTRATGRAPGPEGEVEIAYEAAAGDSIDQLAVGVAAPDLFGEGELEPVLG 1517  
QY 1013 GEGAGRVAVGEGNGVGVOPVATASAGAFAHTVTSATLVLPKPAISATEAAMPVA 1072  
| : : : : :  
Db 1518 SEBAGRVAVGEGNGVGVOPVATLAAGVAFHTVTSATLVLPKPAISATEAAMPVA 1577  
QY 1073 YLTAWVALDRARLQPERVLIHAATGVGLAAVQMAQHAQVAGVHATAGTPEKRAYLESL 1132  
| : : : : :  
Db 1578 YLTAWVALDKVAHQAQGERVLIHAAGGIGLCARVMAQORGAQVAYADDPPEKRAYLESL 1637  
QY 1133 GVRVSDRSRDRFVADYAVATGEGVDVINSLSGELIDKSFNLSRSHGFVFLGKRDY 1192  
| : : : : :  
Db 1638 GVRVSDRSRGRFVADYAVATGEGVDVINSLSGELIDKSFNLSRSHGFVFLGKRDY 1697  
QY 1193 ADNOLGRPFLRLNLSFVLDLRGMLEKRPVRLLELGLTAAGV----- 1239  
| : : : : :  
Db 1698 ADTQPGPLPLLRNFSFQVLDLRGMMDQPARIRALDELGLVAAAGAISPLSGGLRVGSS 1757  
QY 1240 FTPEPPIALPIARVADAFRSMAQAHLKVLTLGDEVOAIRPTHA----- 1286  
| : : : : :  
Db 1758 LTPPEPVTFFISRAAEFRMAQOHLGKVLTLIDPEVAVIRAPAESVAVRADGTYLT 1817  
QY 1287 -----GAG----- 1289  
| : : : : :  
Db 1818 GGLGGLGLRVAGWLAERAGQLVLVGRSAAAEQRAAVALAENHARVTAKADVADR 1877  
QY 1290 ----- 1289  
| : : : : :  
Db 1878 QIERVLEVTASGMPLRGVVHAAGLVDDGLMOQTPARLTWGPVKVQALHLLTREA 1937  
QY 1290 ----- 1289  
| : : : : :  
Db 1938 PLSPFVLVYASAGLFGSPQGNVAAANAFDLSLHHRAGLPALSTIDMGTFEIVGAVA 1997  
QY 1290 -----PST----- 1292  
| : : : : :  
Db 1998 QENRGARLISGRKGITPDEGLSALARLLLEGDRVQGVIPITPQWVEFYATASRRLS 2057  
QY 1293 -----GDRDLIDRLASAPARAAALAEFLTOVSQVARTREIKVGAHALF 1338  
| : : : : :  
Db 2058 RLVTTOQAVADRAAGDLDLEQLASAPASAPAGLLQVAVVQVSHVRLBEDKIEVDAPL 2117  
QY 1339 TRLGMDSIMAVELNRIEASLKLSTFLSTSPNIALLAQNLL-DALATATL 1389  
| : : : : :  
Db 2118 SSMGMDSIMSELNRIEALGVAAPALGWTYPTVAITRWLLDMLAVRL 2169

RESULT 5  
ID OQK127 PRELIMINARY; PRT; 7257 AA.  
AC OQK127;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE EPOD.  
GN EPOD.  
OS Polyanthum cellulosum.  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
OC Soranginae; Polyangiaceae; Polyangium.  
OX NCBI\_Taxid=56;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SMP44;  
RX MEDLINE=20293058; PubMed=10831849;  
RA Julien B., Shah S., Ziemann R., Goldman R., Katz L., Khosla C.,  
RT "Isolation and characterization of the epoIII gene cluster  
cluster from Sorangium cellulosum.";

RL Gene 249:153-160(2000).  
[2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=SMP44;  
RX MEDLINE=20115953; PubMed=10649995;  
RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.,  
RT "Cloning and heterologous expression of the epoIII gene cluster.";  
Science 287:640-642(2000).  
CC - SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES  
(SDR) FAMILY.  
DR EMBL: AF217189; AAF62883.1; -.  
DR HSSP: P25715; IMLA.  
DR InterPro: IPR001227; Ac transferase.  
DR InterPro: IPR002198; ADH short.  
DR InterPro: IPR002085; Adh zn family.  
DR InterPro: IPR004410; FAD.  
DR InterPro: IPR000794; ketoacyl-synt.  
DR InterPro: IPR000209; peptidase\_58.  
DR InterPro: IPR006162; Pantoic attach.  
DR InterPro: IPR006163; Pp bind.  
DR InterPro: IPR002364; OOR zeta crystal.  
DR Pfam: PF00658; Acyl transferase.  
DR Pfam: PF00106; adh short; 1.  
DR Pfam: PF00107; adh zinc; 2.  
DR Pfam: PF00109; ketoacyl-synt; 4.  
DR Pfam: PF02801; ketoacyl-synt C; 4.  
DR Pfam: PF00550; pp-binding; 4.  
DR TIGRfam: TIGR00128; fad; 4.  
DR PROSITE: PS00075; ACP DOMAIN; 4.  
DR PROSITE: PS00606; B KETOACYL SYNTHASE; 4.  
DR PROSITE: PS00012; PHOSPHOPANTHEINE; 2.  
DR PROSITE: PS01162; OOR ZETA CRYSTAL; 1.  
DR PROSITE: PS00136; SUBTILASE Asp; 1.  
KW Oxidoreductase; phosphopantetheine; Transferase.  
SQ SEQUENCE 7257 AA; 764877 MW; 3A53650FAD2CEB9 CRC64;

Query Match 46.1%; Score 3322.5; DB 2; Length 7257;  
Best Local Similarity 41.2%; Pred. No. 1.5e-153;  
Matches 766; Conservative 161; Mismatches 431; Indels 531; Gaps 25;

QY 9 AEDPIATVAGSGRLPGVYIDLSGFVTLLEGSRDVTGRVPAERWDAAMFDDPDPAGKT 68  
| : : : : :  
Db 3021 ASDPFAIVAGACRFGGVEDLESYQWQLAEGVVSABVPARMDADYDDPEIPGRGT 3080  
QY 69 PVRASPELDVACFDSFPGISPREALRMDPARLLEVCWEALNAALAPSLVGTENG 128  
| : : : : :  
Db 3081 YVTKGAFLELDLQRLDTPFRISPREMAGSLDPOQRLLEVSWEALBSAGIAPDLRPSPTG 3140  
QY 129 VFIGIPSEYEALPOTASAEIDAAGLGTMPVSGAGRIAYALGLRCPVAVDTAYSSS 188  
| : : : : :  
Db 3141 VFGAGAPNEYVYQRLGFLDGAAGLVGGTGNLVSVAAGRLSFLGLHGPTLAMDTACSSS 3200  
QY 189 LVAVHLACQSLRSGECSTLAGVSLMSPSTLWLSKTRALARDRCGAFSAEADGFR 248  
| : : : : :  
Db 3201 LVALHLACQSLRSGECQALVGVGNVLLAPETFFVLRSRRALSPDRCFTFSADAGYAR 3260  
QY 249 GEGCAVAVLKRISGARADDDRIATVRSAINHDGASGLTVPNGSSOEIVKRALADG 308  
| : : : : :  
Db 3261 GEGCAVAVLKRISGARADDDRIATVRSAINHDGASGLTVPNGSAQALRLQALRQALSOAG 3320  
QY 309 CAASGVYEAAGTGTLLDPIETIQALNAVYGLGRDVAATPLLIGSVKTLGHEVYASGT 368  
| : : : : :  
Db 3321 VSPVDVDFEHCCTGTALDPIEVQALSEYVGGREDEDRPLVIGAVKAVAHLEAASGLA 3380  
QY 369 GLIKVVLSTOHGOIPAHHLAQAALNPRIISGDLRLTYTRARTPMPNTTRRAGVSSFGMS 428  
| : : : : :  
Db 3381 SLIKAVIALRHEQIPAOPEIGELNPHLPNNITLPAVAPRAVPGRCARPRRAGVSAFGIS 3440  
QY 429 GTNAHVLEAPAPATCTPPAPRPAELVLSARTASALDAQARLDHLETYSQCLGCV 488  
| : : : : :  
Db 3441 GTNAHVLEAPAPATCTPPAPRPAELVLSARTASALDAQARLDHLETYSQCLGCV 500  
QY 489 AFSIATTRAMEHRLVAATSREGRLAALAAQOGTSPGAVRSIADSSRGKLAFLPTQ 548



```

Db 3501 ASLSTRSPMEHRLAIATTSREALRGALDAAQOTPOGAVGKGVSSRGKLAFLFTQ 3560
Qy 549 GAOTLMGSGLYDVMSAPREAPDLCYRLFNOELDRFREVMMAEPASVDAAILDQAFPO 608
Db 3561 GAOIMGSGLYEAMWAFREAFDRCAVLDRELDQFLREVMMAEPALQARLDQAFVQ 3620
Qy 609 PALFTFEVALAALMRSWGVEBELVAGHSIGELVAAACVAGFSLIEDAVFLVAARGRLMAL 668
Db 3621 PALFTFEVALAALMRSWGVEBELVAGHSIGELVAAACVAGFSLIEDAVFLVAARGRLMAL 3680
Qy 669 PGGAMVSTLEAPADVAAVAPPAASVSTAAVNAPOQVVIAGAGQVHAIAAAMARGAR 728
Db 3681 PGGAMVSTLEAPADVAAVAPPAASVSTAAVNAPOQVVIAGAGQVHAIAAAMARGAR 3740
Qy 729 TAAHSHAFHSPLMAPMLEAFGRVAVESYRPSITLVSNLSGKCTDEBSPGVMVH 788
Db 3741 TAAHSHAFHSPLMAPMLEAFGRVAVESYRPSITLVSNLSGKCTDEBSPGVMVH 3800
Qy 789 AAEVVRFADGVYALHAAGAGTEVEVGPKSTLLGLVPACMPDARPALASSRAGRDEPATV 848
Db 3801 VASAVRFGDGAKALHAAGATFVEIGPKVLLGLLPACIGADAVLVPRLRDRSCFV 3860
Qy 849 LEALGGLMVGGLVSMAGLPESGGRVPLPTYPQOERYWIDTKADA----- 896
Db 3861 LEALGGLMVGGLVSMAGLPESGGRVPLPTYPQOERYWIDTKADA----- 3920
Qy 897 ----- 896
Db 3921 GVGCLMPCGVLAHNVLSIGRRHOPFLGDHLVFGKVVPAGFHAVALISIAERWPERAIEL 3980
Qy 897 ----- 896
Db 897 ----- 902
Qy 3981 TGVFELKATAMEBDEVELHAVLTPREAGDGYLFELATILAPETERKMTTHARGVQPTD 4040
Qy 903 -APGA----- 906
Db 4041 GAPGALPRLVLEDBRAIQPLDFAGFLRLSAVRIGWPLMRLQDGRVGDASLATLVPT 4100
Qy 907 ---GHD-----EVEGGA-----VR-GGD 921
Db 4101 YENADVAPLHPLILDNGFVALLATRSBEDDGTPLPFAVERVMWRAPVGRVRCGV 4160
Qy 922 RRSAA-----RLDHPP-- 932
Db 4161 PRSQAFGVSSFVLVDTEGEVVAEVEGFCRABREVFILROESGASTAALYRLDWEPAIPL 4220
Qy 933 -----ESGR 936
Db 4221 DAPARMEBSWVVAAPGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVTCIMEPGA 4280
Qy 937 REKVAAAGR----- 946
Db 4281 HEEAPRAAQVATEGLSVQALRDRAVRLMWTTGAVAVEAGERVQVATAPVWGLGRVW 4340
Qy 947 ----- 946
Db 4341 QERPELCTVLDEPEVDARSADVRLRELGADETOVVPFSGERRVAVLKATTPBCL 4400
Qy 947 -----PFRLEIDPEGVLDHVLVYTERRAPGLGEVEIANDAGLSFNDVOLALGNVDD 1000
Db 4401 LVPDAESYRLTAGQKGTLDQLRLAPQRORAPGGEVEIKVTASGLNFRVLVAVLGM---- 4456
Qy 1001 LBGKNPRLLLGEGEGRIVAAGEGNGLVNGQPVIALSAGFATHYVTSALVLPKPPQA 1060
Db 4457 YFGDAGP---MGDDAGVITAVGQGVHNSVSDAVWTL--GLHLRFVYDARLVVRQBPAG 4511
Qy 1061 LSAIEAAMPVAYVLTAWVALDRIARLQPERVLIHAATGSGVLAIVQWQAHGAEVHATA 1120
Db 4512 LTPQAQATVPVAFITMLLHLGNRGERVLIHAAAGVGMAAVQIRWIGAEVAFATA 4571
Qy 1121 GTPERKAVYESLGV--RIVSDRSRDFVADVRAWTGEGVDVVLNSLSEGLIDKSFNLIR 1178

```

```

Db 4572 -SPSKAAVQAMGVERTHIASRTLEFAETPRQVTCGRGVVDVVLNALAGEFVDASLSLT 4630
Qy 1179 SHGRFVEIGKRCVADNQ--LGLRPFLNLSPLVDLGMVLEPARVRLAEELGLIA 1236
Db 4631 TIGRFLNGKTD-IDBRAVAAAHFGVRYFDLIEL-----ADRTREILERVVEGFA 4683
Qy 1237 AGVTPPPPIATLPIARVADAFRSMAOAOHLGLVLTLDPEVOIRIPT-----HAGAP- 1290
Db 4684 AGLHALPVAHNAIKAEAPFPMQARHOGKVVL-LRAPSAPLAPGTIVLLTGGLAL 4742
Qy 1291 -----STGDRDLRLASAPARAALAEFLTQVSQVLRTEPIKVG 1333
Db 4743 GLHVARWLAQOQAPHVLTGKRG-LDTPGAAKAVAEIHALGARVTIAASDVADRNALAEV 4801
Qy 1334 AEAL-----FTFLGMSLMAVBLRRIRIESLTKXUST 1365
Db 4802 LQATPAEMPLQGVHIAAGALDDGVLDDEOTTRFSSVLAIPKVTGANVLEHLLTGNDLAFV 4861
Qy 1366 TFLSTS-----DNIALLAQNLIDALATALSLEVAENLRAGVOND 1406
Db 4862 LESSMSGLLGSAQGSNYA-AANTFLDALAHRRAEGLAQAOSLAWKPMWD 4909

RESULT 6
Q9L8C7 PRELIMINARY; PRT: 7257 AA.
ID Q9L8C7;
AC Q9L8C7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Polypeptide synthase.
GN EPOC.
OS Polyanium cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Soranginaeae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=So ce90;
RX MEDLINE=20130945; PubMed=10662695;
RA Molnar I., Schnupp T., Ono M., Zirkle R.E., Milanow M.,
RA Nowak-Thompson B., Engel N., Toupet C., Stritzmann A., Cyr D.D.,
RA Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT epothilones A and B from Sorangium cellulosum So ce90.";
RL Chem. Biol. 7:97-109(2000).
DR EMBL: AF210843; AAF26921.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR007094; ketoacyl-synt.
DR InterPro; IPR006162; peptidase_S8.
DR InterPro; IPR006163; pp_bind.
DR InterPro; IPR002364; QOR_zeta_crystal.
DR Pfam; PF00698; Acyl_transf; 4.
DR Pfam; PF00107; adh_zinc; 2.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.
DR TIGRFAMs; TIGR00128; fabD; 4.
DR PROSITE; PS50075; ACP_DOMAIN; 4.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
DR PROSITE; PS00162; QOR_ZETA_CRYSTAL; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 7257 AA; 764272 MW; 621A6820FA57ACE9 CRC64;

Query Match 45.9%; Score 3311.5; DB 2; Length 7257;
Best Local Similarity 41.1%; Pred. No. 5.2e-153;
Matches 784; Conservative 162; Mismatches 433; Indels 531; Gaps 25;

```

QY 9 AAEPIAIVGASGRLEPGVITLDSGFWTLEGSRDITGVPAERMDAAAFDPDPA BGT 68  
 Db 3021 ASBEPPIAIVGASGRLEPGVITLDSGFWTLEGSRDITGVPAERMDAAAFDPDPA BGT 3080  
 QY 69 PUTRASFLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIAPSAVLGTETG 128  
 Db 3081 YVTKAFILRLDQGLDATTFRISPREAMSLDPQGRLLLEVSWALBSAGIAPDTRLRSPGT 3140  
 QY 129 VFIIGTSEYEALPQATSAEIDAGGLTSPSVGAGISYALGLRGPCVAVDTAYSS 188  
 Db 3141 VFVGAAPNEYTQRLRFTDGAAGLYGGTNMLSVTAGRLSFTGLGPTLADITACSS 3200  
 QY 189 LVVVLHACOSLRSGESTALAGVSLMSPSTLWLSKTRALARDCKAFSAADQFGR 248  
 Db 3201 LVVVLHACOSLRSGESTALAGVSLMSPSTLWLSKTRALARDCKAFSAADQFGR 3260  
 QY 249 GEGCAVVLKRLSGARADGRIIAVIRGSAINHDSAGSLTPNGSSQELVKRALADAG 308  
 Db 3261 GEGCAVVLKRLSGARADGRIIAVIRGSAINHDSAGSLTPNGSSQELVKRALADAG 3320  
 QY 309 CAASSYGVYEAHGTGTTGPIELQALNANYGLGRVATPLLGSVNTNIGHEVYASGIT 368  
 Db 3321 VSEVDVDFVECHGTGALGDIIEVQALSEVYGGSGDRPLVGAAGANVAHLEAASGLA 3380  
 QY 369 GLKLVVLSLQHGQIPAHILHAQALNPRISMGDLRLTVTRARTPMPDMNTPRAGVSSFGMS 428  
 Db 3381 SLIKAVIALRHEQIPROPELGEINPHLPMNTLPVAVRKAVPMGRKARPRAGVSARGLS 3440  
 QY 429 GTNAHVLEEAAPACTTPRAPERPAELVLISARTASALDAQAARLRLDLETYSQCLGV 488  
 Db 3441 GTNAHVLEEAAPACTTPRAPERPAELVLISARTASALDAQAARLRLDLETYSQCLGV 3500  
 QY 489 ABEIATTRSMERILAAVATSRREGRLAALAAAGQTSPPANVSIASSSGKLAFLFTGQ 548  
 Db 3501 ABEIATTRSMERILAAVATSRREGRLAALAAAGQTSPPANVSIASSSGKLAFLFTGQ 3560  
 QY 549 GAOTLGMGRGLYDWSAFREAFDLCVTLFNOELDRPREVMAEPASVDAALLDQTAFTQ 608  
 Db 3561 GAOTLGMGRGLYDWSAFREAFDLCVTLFNOELDRPREVMAEPASVDAALLDQTAFTQ 3620  
 QY 609 PALFTEYALAAALMRSGVEPELVAGHSIGELVAAVAGVSEDAVFLVARGRLMOAL 668  
 Db 3621 PALFTEYALAAALMRSGVEPELVAGHSIGELVAAVAGVSEDAVFLVARGRLMOAL 3680  
 QY 669 PAGAMVSIAPREDAVAAVAPHAASVSIANAAPDOVITAGACOPHAIILAAAMARGAR 728  
 Db 3681 PAGAMVSIAPREDAVAAVAPHAASVSIANAAPDOVITAGACOPHAIILAAAMARGAR 3740  
 QY 729 TKALHVSHPPLMAFMLBAFGRVSAESVYRRPSIVLVENLSGKACTDEVSFGYVVRH 788  
 Db 3741 TKALHVSHPPLMAFMLBAFGRVSAESVYRRPSIVLVENLSGKACTDEVSFGYVVRH 3800  
 QY 789 AREVVRADGVKALHAAGAGTVEVGPKSTLLGLVPCMPDARALLASSRAGDEPATV 848  
 Db 3801 AREVVRADGVKALHAAGAGTVEVGPKSTLLGLVPCMPDARALLASSRAGDEPATV 3860  
 QY 849 LELGLGMAYGLVSWAGLFRSGGRVPLPTVMQRRYMTDTAADA 896  
 Db 3861 LELGLGMAYGLVSWAGLFRSGGRVPLPTVMQRRYMTDTAADA 3920  
 QY 897 896  
 Db 3921 GVGICMGAVLHVLSIGPRHOPFLGHLVFGKVVGCAHVAVILSIAERWERAIEL 3980  
 QY 897 902  
 Db 3981 TGVEFLKAIAMPDQVEVLAHLVTPBAAGGYLFEIATLAETERMTTHARGVQPTD 4040  
 QY 903 -ARCA 906  
 Db 4041 GARGALPRLLEVEDRAIOPIDFAGFLDRLSAVRIGWPLRMLODGRVGDASIALTLVPT 4100

QY 907 ---GHD-----EVEBGA-----VR-GSD 921  
 Db 4101 YRNAHVAPLHLILDNGFAVSLSTRSEPPEDGTPPLFAVERVMWRAPVGRVRCGV 4160  
 QY 922 RNSA-----RDLHPP 932  
 Db 4161 PPSQAFVSSPVLVDETEGVVAEVEGFCRARPREVFLRQESGASTAALYRLDWEAPAP 4220  
 QY 933 -----ESGR 936  
 Db 4221 DAPAEIESKVVVAAPGSEMAAALATRLNRCVLAEPKLEALLAGVSPAGVICTMEAGA 4280  
 QY 937 REKVEAAGR----- 946  
 Db 4281 HEEAPAAARVATEGLSVVQALRDRAVRLMWTVGAVAEAGBRVQATAPVWGLGRVYM 4340  
 QY 947 ----- 946  
 Db 4341 QERPELCTLVLEPEADARSADVLLRELGRADDETOVAFRSGRKRVAVKATTPBGL 4400  
 QY 947 -----PRLTIDEPGLDHLVLRVTERAPGLGEVEIADVDAAGSFNDVOLALGMVPPD 1000  
 Db 4401 LVPDAESTYLEAGQKTDQLAPAKORAPGEGVEIKVTASGLNFRIVLAVLGM----- 4456  
 QY 1001 LPCKPNPULLGEGECAGRIAVAGEVNGLVGQPVIALSAGAPATHVTTSAALVLPFQA 1060  
 Db 4457 YPEDDAGR---MGDDCAGVATAVGGQVRHVAVDGAVVTL---GTLHFRVTVDAELVVRQAG 4511  
 QY 1061 LSAIEAAMPVAVLTWVYALDLRIARLOBERVLINAAATGUGLAANVQMAOHGAEVHAHA 1120  
 Db 4512 LTPQAATVPVAFLTMLMLHDGLNLRGERVLIHAAAGGVMAAVQIARWIGAEVFATA 4571  
 QY 1121 GPBEKRAVLESIGV---RYVSDRSDFVADVAVMTGEGEDVVLNSLSGELIDKSFNLIR 1178  
 Db 4572 -SPSKMAAVQANGVPRTHIASRTLEFAETFRQVTCGRVDVVLNAAELFPDASISLIS 4630  
 QY 1179 SHGRFVELGKRCYADNO---LGLRPELRNLSFSLVLRGMLEPRPARVALLLEELIGLIA 1236  
 Db 4631 TGRFLEMGKTD-IRBRAVAAAHPCGRVYRFDIEL-----APDRTREILERVEGPA 4683  
 QY 1237 AGVTFPPPIATLPIARVADAFRSMQOHLGKLVLTGPEVOIRPT-----HAGAGP- 1290  
 Db 4684 AGHTRALPVHAFATTEAAAFRMAQARHQGVVL-LPPPSAAPLAPFTVLLTGGLGAL 4742  
 QY 1291 -----STGDRDLDRLASAAPARAALAEFLRTQVSGVLRTPETIKVG 1333  
 Db 4743 GLHVAWLAQGVPRHVLTRRG-LDTPGAARVAEIEALGARVTTAASDVADRNALAEV 4801  
 QY 1334 AEARL-----FTRLGMSLMAVELRNRIEASLKJLKJST 1365  
 Db 4802 LQIIPAEWPLQGVIIHAAGALDQVLEQTTDRFSRVLAPEKVTGAMWLHELTAGNDLAFV 4861  
 QY 1366 TPLSTS-----PVALLAQNLDAATALSLEVAENLRAGVQND 1406  
 Db 4862 LFSMSGGLLSAGQSNYA--AAVTFLDALAAHRAAREGLAASLAWGWSMD 4909  
 QY 1406

RESULT 7  
 QBRJY1 ID QBRJY1 PRELIMINARY; PRT; 2218 AA.  
 AC QBRJY1;  
 DT 01-JUN-2002 (T-REMBLrel. 21, Created)  
 DT 01-JUN-2002 (T-REMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (T-REMBLrel. 22, Last annotation update)  
 DE Scif. protein.  
 GN Scif.  
 OS Stigmatella aurantiaca.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 OC Cystobacterineae; Cystobacteraceae; Stigmatella.  
 NCBI\_TaxID=41;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5g a15;

RA Gaitatzis N., Siłakowski B., Kunze B., Nordstiek G., Blocker H.,  
 RA Hofle G., Muller R.;  
 RT "The biosynthesis of the aromatic myxobacterial electron transport  
 RT inhibitor stigmatellin is directed by a novel type of modular  
 RT polyketide synthase";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A4421825; CAD19090.1; -  
 DR InterPro: IPR001227; Ac transferase.  
 DR InterPro: IPR002085; Adh. zn family.  
 DR InterPro: IPR000794; Ketoacyl-synt.  
 DR InterPro: IPR006163; pp\_bind.  
 DR Pfam: PF00698; Acyl-transfer; 1.  
 DR Pfam: PF00107; adh\_zinc; 1.  
 DR Pfam: PF00109; ketoacyl-synt; 1.  
 DR Pfam: PF02801; ketoacyl-synt; C; 1.  
 DR Pfam: PF00550; pp-binding; 1; C; 1.  
 DR PROSITE: PS50075; ACP DOMAIN; 1.  
 DR PROSITE: PS00606; B\_KETOACYL SYNTHASE; 1.  
 DR Phosphotransferase; Transferase.  
 KW SEQUENCE 2218 AA; 23739 MW; B85339D5457PA4F5 CRC64;

Query Match 36.5%; Score 2634; DB 2; Length 2218;  
 Best Local Similarity 35.4%; Pred. No. 1.7e-120;  
 Matches 642; Conservative 182; Mismatches 447; Indels 542; Gaps 25;

QY 8 RAADPPIAVGASCRPLPGVITLDSGFWTLLEGSDTVGRVPAERMDAAAFDPDPAPGK 67  
 DB 24 RLAPPEIATVGMACRIPGVSVDSDSYNDLLKGGDVAITEIPAAKMDINAFYDNDSPK 83  
 QY 68 TTVTRASFLSDVACFDASFFGISPREALRMDPAHLLLEVCWALEMNAIABSALVTET 127  
 DB 84 TTRWCGFLDVEGFPDHLFGISPREASMDPHQRLFLEVAMEALEMNAQOQDRLAGSKT 143  
 QY 128 GVFITGIGSEBEYE-AALPOATASAEITAGGCTGMEVAGRISYALGLRPPCAVDATYS 186  
 DB 144 GVFVGLTSDYSPVVLSDNA--INVYGTGTANNVAAGRSLTYLLDLOGPSFVVDTACS 200  
 QY 187 SSLVVAHLACOSILRSGECSTALAGVSLMSPETLVMLKTRALARDGCKAFSAEADP 246  
 DB 201 SSLVVAHLACOSILRSGECSTALAGVSLMSPETLVMLKTRALARDGCKAFSAEADP 260  
 QY 247 GRGEGCAVVVLKRLSGARDGRILAVIRGSAINHDSAGSLTPVNGSQEIVLKRLALD 306  
 DB 261 ANGEGCGVILKRLSDPADGRILAVIRKASAINQDGRITGLTAFSGAQVLLRQALTR 320  
 QY 307 ACCAASSVGVENAHGCTTLGDPITLQALNAVYGLGRDVAITPLISVYNTNGHEPYAG 366  
 DB 321 AGISPSQVTVYEAHGTGTPLDPIEVEALAEVYGOFRPSGPCVLGSVXANLGHGTGAAG 380  
 QY 367 ITGLKLVVLSLOHGOI PAHLHAQALNPRISWGDRLRTVTRARTPMPDMNTPRRAGVSFG 426  
 DB 381 VAGLIVVLSWQNEALIPOLHYRELNSISLENTSLVITPELHPWAGCGPRIAGISAG 440  
 QY 427 MSGTNAHVLEENAPATCTPPAPERPAELVLISARTASALDAQAPALRDLHLETYPSCQ 486  
 DB 441 MSGTNAHVLEENAPATCTPPAPERPAELVLISARTASALDAQAPALRDLHLETYPSCQ 500  
 QY 487 DVAFSLATTRSAMERFLAVATSRBGLRALDLAAAGQTSPGAIVSIADSRCK-LATLF 545  
 DB 501 DLYCSANTGMSKFEHLIAPAGTTAQLSERLSAFLEGGQPOAGTSYQVTVGRGPGVAFIF 560  
 QY 546 TGGGATGLMGRGLVYVMSAFREAPDLCVRLFNQELDRPLREVMMAEPASVDAALLDQTA 605  
 DB 561 TGGGATGLMGRGLVYVMSAFREAPDLCVRLFNQELDRPLREVMMAEPASVDAALLDQTA 618  
 QY 606 FTQPALFTTEYALAAALMRMGVEPELVASHISIGELVAAVAGVSLIEDAVFLVAAGRLM 665  
 DB 619 FTQPALFAEWALAEIMRSMGVPTVVMGSHVGEYVAAACVAGFTLEBGLALIVERARLM 678  
 QY 666 QALPRAGANVSI EAPADVAALVAAPHAASVSI AA VAPDOVYTAGGQVYHAI AAMAR 725  
 DB 679 QSLPRGEMAAVFTDGARTAOAIAPYASOVSI AAFNGPSETVYSGDGKAVEAILGALSAE 738

QY 726 GARTALVSHAFHSPLMAPMLEAFGRVAESVYRPSIVLVNSLGNACTDEVS PGW 785  
 DB 729 GVKVIRLIGSHAFHSPLMDPMDAFERAAKTRFMPKITLISNLG-GPSEFSAAL- 796  
 QY 786 VHAHEVVRFPADGVKALHAAGATVEVGPXSTLLGLVPACMPDARPALASSRGRDP 845  
 DB 797 RHAREPVRFPDGMALARGVSVAVEIGPATLLGISRCIPESVAMVPLSRDKDEM 856  
 QY 846 ATVLALGLVMAVGLVSMAGL--FPGSGRVPLPTYPWQBERY----- 887  
 DB 857 EVLLATLGLHLSVGRVPVWAGFSDYPR--RRVSLPNTPFQHERALKOMDGRGVAIPSGQ 914  
 QY 888 ----- 887  
 DB 915 VVGHPLLSRLRVATVDGTLFESKFRADBPSPFSFHRVYGMSTVPATVLEGALAAEBI 974  
 QY 888 ----- 887  
 DB 975 FGAGHALESVDIOEALVLSDAPTVOYLIGSREBGTARFOVFLASSGSGEGWTTTH 1034  
 QY 888 ----- 890  
 DB 1035 SSGQIRIASGAEPLSGTLPTEQALEAIRTRCPGLEPRAFYDWLEGGLEYGPRFRGQ 1094  
 QY 891 ----- 912  
 DB 1095 QVMRGDGBALGLVELPESAVDEAKAFKVNPLGLMDACVQLFGVIEYREGAQDTEVYLPVS 1154  
 QY 913 ----- 914  
 DB 1155 IDRYLKGRLGSKVSHARIRPEEGTARETLKADIRLYDEAGRIVAEEGMCIKRAPREM 1214  
 QY 915 ----- 928  
 DB 1215 IARFQAOFSMLHEWEKKEVRSAAAGARGSGRWVAFDRSGVAEALIDRLTAGRP 1274  
 QY 929 ----- 936  
 DB 1275 VVRVLADHYSREGDVLRLHPQEPHVRSLMESAASGGAIVTRYLYWGLDSSGVSNGE 1334  
 QY 937 REKVE--AAGDRP----- 947  
 DB 1335 ASRVDTRAIGSNPLHLVQELAGQGSQHLFLVTRGAQATTREHAPIDAFQASLWGFAAT 1394  
 QY 948 ----- 947  
 DB 1395 IAFHEPRLPVRIIDLEPSRKGSPRETEILLEGVTRSDAASADEAIARGLISRLVSLMR 1454  
 QY 948 ----- 989  
 DB 1455 QAKRASNRKLRIPAPAPRLEISQRTLEDTLVAPVRRPPGPEVEIRVAPATGLNPRD 1514  
 QY 990 VQALGMVPDDLPGKPNPPLLIGECACAGRIVAVGCVGLVVGQVIA-LSAGAPATHVT 1048  
 DB 1515 VLNALGMVPGD-----FGLLGECACAGVISAIGVGVGFKVGDVVVAIVISGFATYCT 1567  
 QY 1049 TSAALVLRPPQALSAIEAAMPVAVLTAMVYALDRARLQPGERVLIHAATGCVGLAAQW 1108  
 DB 1568 VLADVVAHPPALISPAOAAAIPIAELTAOYGLQOGRMAAGSRVLIHAAGGVGMAAVAL 1627  
 QY 1109 AQHVGAETHATAGTEPEKAVYLESLGRVYVSDRSRFPADVAVAMTGGEGVDVNLSSGE 1168  
 DB 1628 AKRAGAEVPAITSS-GKNDVLEAMKEVERKMSRTLDFADEVWARKREGGIDIVLMSLAD 1686  
 QY 1169 LIDKSFNLIRSHGRFVELKGD CYADNQLGLRPLRLNLSFSLVDLRGMLEPARVAL 1228  
 DB 1687 FIPKSLVLSRSGRFLIEIGKGVMTSAQVAER--YPGVSVAIYDGBVGTQPSIIRSMF 1744  
 QY 1229 BEILGLIAAGVTPPPIATLPIARVADAFRMAQOHGKVLTLGD--PEVQIITPTHA 1286  
 DB 1745 RALMFEFGTLVPPPLLVFALQDAVEAFRHMAALAKHGVKIVITVNDGAP-SAETTARA 1803  
 QY 1287 GAG-PTGDRDLL 1298



```

QY 949 -----RLEIDEPGVLDHLVL-RTERRAP-GLGEVEIAVDAAGLSFNDV 990
DB 3464 NARSDIELIAPDASNMRLHIPTKGNFDLTLVDAPLAPRLHGVRAVHAANFNDV 3523
QY 991 QALCMVDDLPKKNPRLGCEQAGRIIVAVEGVNGLVQPIYALSAGAFATHTTS 1050
DB 3524 LDTLGL-----YFGDAGP---LOGEGAGIVTEVGPVSRTYVGDVWVGIFGAACGPTALAD 3576
QY 1051 AALVLRPPOALSAIEAAMPVAYLTAWYALDRIBLQGERVLIHAAGCGVGLAVQNAQ 1110
DB 3577 ARMTCPPIPIANSFAQASVPITLYAYGLVDLGHKKNQVLIHAAGGVGTAAVQIAR 3636
QY 1111 HVGAENVHATAGTPEKRAVYLESIGV--RYVSDRSDFVADVAWMTGEGEVVDVLSIGE 1168
DB 3637 HLGAVVPNTA-SAGKMSALRALGFPDHAHLASSRDIDFEGHFLRSTHGKGVDDVLDCLARE 3695
QY 1169 LIDKSFNLIRSHGRVVEIGKRDQVADNQLGLR-PLRLNLSFSLVDLRGMKLEPRARVAL 1227
DB 3696 FVDASLRIMPSCGSEFVEMGKTDIRREPDVAVGAVPGVVRAFDLIE-----AGPDRIEQM 3749
QY 1228 LEEELGLIAGVFTPEPIATLPIARVADAFRSMAQOHGKVLTLG---DEVOQIRIPT 1264
DB 3750 LAELSLTEBGRALRPPITSMDIRHAPQAFRALADRRHVKFVLTIPRIDEGLVLI-- 3807
QY 1285 HAGAPSTGDRDLDRLASAAPARAALAEFLRTQVSQVLTPEIKVGAELFTRLGMD 1344
DB 3808 -----TGCTGTALVARNHLVARGHAKHLLTSR--QGANAP-----GAEN--SRTEBLE 3852
QY 1345 SLMA-VEIR--NRISASLKLKSTFTLSTSPNIAL--AQNILDALATLSIERV 1394
DB 3853 ALGASVTLRACDAADPRALQALLDSIPSNHPLTAVVAAGALDDGLGMSPERI 3907

```

## RESULT 9

```

Q8GBX6 PRELIMINARY; PRT; 5017 AA.
AC Q8GBX6;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Polypeptide synthase.
GN SPH.
OS Polyangium cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangelineae; Polyangiaceae; Polyangium.
CX NCBI_Taxid=56;
RN [1]
RP SEQUENCE FROM N.A.
RA Knauber J., Bioecker H., Mueller R., Nordstiek G., Beyer S.;
RT "The spirangeliene synthase from Sorangium cellulosum So c690 - Module
RT duplications as strategy behind the evolution of polypeptide
RL synthases.";
RL Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ505006; CAD3449.1; -.
SQ SEQUENCE 5017 AA; 52555 MW; 08BED56182669493 CRC64;

```

Query Match 33.5%; Score 2416.5; DB 2; Length 5017;  
 Best Local Similarity 39.8%; Pred. No. 2.4e-109;  
 Matches 639; Conservative 186; Mismatches 521; Indels 259; Gaps 41;

```

QY 9 AAEDEPIAVGASCRPLGCVITDLSGFWTLLEGSRDTVGVAPARMWDAAMFDDPDPAQKT 68
DB 1586 ASDEPIAIVGVGLRPLGVDVDDALFRLLGGERDAVAPIPASRMVGVAVFDDPEAKXT 1645
QY 69 PYTRASFLSDVACPDASFFGISPREALMDPAHRLLEVCWALENAAIAPSAVGTETG 128
DB 1646 YVRHAMLDKRYDLDFAGFFGISPREARHYDPQHRLLLETAAQTAATVPSLSDSTTG 1705
QY 129 VFIGIGPEVEYALQATASAEIDAAGGLTMRPSGAGRISALGRPCVAVDTAYSSS 188
DB 1706 VFGVAGASDY--AVLQSSAE--DAEAYAAWGTAASFAAGRLATLGLQGPALSIDTACSSS 1762

```

```

QY 189 LVAVHLAGQSLRSECESTALAGVSLMLSPSTLYMLSKTRALARDGRCKAFSAEADGFR 248
DB 1763 LVALHLAQSLRQCECDLALAAVGVMAFPEVFIILSTRALAPDGRKTSANADGFR 1822
QY 249 GEGCAVVLKRLSGARADGRILAVIRGSAINHGGASGLTPVNGSSGEIYLKRLADRG 308
DB 1823 GEGVAVLLELRBARANGHFLVAVRGTAANHGGASGIRAPNGSSQOKYLRALDAR 1882
QY 309 CAASGVVEAHGTTGLGDIETIOALNAVYGLRDVATPLIGSVKTNLGHPEVASGIT 368
DB 1883 LAPVDVVECHGIGTISLGDPIEVOALNAVYGEGRAPRRLGLAVKTNIGHLEASGLA 1942
QY 369 GLKLVVLSLQGOIPAHLLAQAALNPRLSMGDLRLVTTRARTPMDW--NTPRAGVSSFG 426
DB 1943 GVVKIVASLRKGTLPATHTNPRNPHIDMDALPRVVDAAAPWESASGSRBRAGVSAFG 2002
QY 427 MSGTNAVVELEAPATCTPPAPERPELL-----VLSARTASALDAQARLRDHE 478
DB 2003 LSGTNAVYVEAP---EPAPRAARGALPLSGALPVLISAKSDAALRAQAALLRDGLA 2058
QY 479 TYPSCIGDVAFSLATTRSAHEHRLAVAATSRREGIRALDAAAGQTSRGAVRSIADSSR 538
DB 2059 RTPDALVDVAALATITRSQFDHRAIYASDHGLVAALBGLAGVAAPGT--VAKGAA 2116
QY 539 GKLAFLFTGQAGQTLGMRGLYDVWSAFREAFDLCVRLFNGELDRPLREVMMAEPASVDA 598
DB 2117 DKLAFLFTGQAGQAAAGRGLYDAFPVPRDALDAVASHLDRELDRLPLRDVLPARGSESLA 2176
QY 599 ALLDQTAFTOPALTFEYALALMRSGVEBELVAGHSIGLVAAACVAGVSLBDAYVLV 658
DB 2177 SLDRTEFTQALPALALEVALFRLVEAMGVTDVLLGHSVGEILAAHNAVAGVSLDACLTV 2236
QY 659 AARGRLMOALPA--GGAWVSIAREADYAAVAAPHAASVIAAVNAPDOVAVIAGQPVHA 717
DB 2237 AARARLMQALPARQGANVTYATATEVEVLAALBEGDGAELIATINAPSTVAGDVAVLR 2296
QY 718 IAAAMAARGARTKALHVSASHFSPMLAPMLEAFGRVASEVSRRPSIVLVNLSGKACTD 777
DB 2297 VAAHFEARGKATRLRSHASHHMDPMLDAFRVABGLFHHPRITIVENVGRLLAAD 2356
QY 778 EVSSFGIVNRHAREVVRPADGVKALHAGACTFEVVGPKSTLGLVPAKMPDAR--PALL 835
DB 2357 EIRSPDYVVRHVRSAVFADQINTLEADGVSSFELGPHGVLSLGSBALNNAQREVAFV 2416
QY 836 ASSRAGREPTVLEALGGLMAVGVLSWAGLF--PSGGRRLPLPTYMPQRRWYMDTAD 894
DB 2417 PALRDGRADVDAITLAALSSLHVRCGRVDMAFPGFPDQIATLPTVAFQDRFMDADA- 2475
QY 895 DAARGDRRAPGAGHD-----EVEEG--GAVVCGDRSARLDH-----PPESGREGX 939
DB 2476 -GATRDEAAPAPBEDDAFWRAVDAGVGL- GATLNASGEBHLGALTLPALSGMRAR 2533
QY 940 VEAAGDRPRLIEDPGLVHLVLRV-----TERRAPGLGEVEIAVDAAGLS--ENDVQ 991
DB 2534 -----DESLVDALRYVWKKPLTPPAADVAGACTWLVTLAAGADALARELE 2580
QY 992 LALGMVPDLPKKNP-----LLGGECAIRIYAVEGVNGLVQPIYALSAGAFATHT 1046
DB 2581 RALTRAGAEVITWPIAPBEADRVRLAVRQCALDAG- LKGIY--SPALDETTLATHT 2636
QY 1047 VTTSA--ALVLRPOALS--AIEAAMPV-----AYLTAV----- 1077
DB 2637 TALRAGLALTLRLOALGDMKIEAPMLVTRGANVSTGRSDRLAAGSMTGCLGRVYGLE 2696
QY 1078 -----VALDRIARL--QPE-----RVLIHA----- 1096
DB 2697 HPERWGLVDVADGDTIDARALDRVLTLASRDEBDOLALRPTGI FARRLVRAVLGDAP 2756
QY 1097 -----ATGVG-----LAAYQAOH--VGAENVHATAGTPEKRAVYLESIGV 1134
DB 2757 ARTYTAGTSLVTGTALGAHVAPPLAGRAEHLVLSRRGEDAPGAGALRAELALGA 2816
QY 1135 RYV-----SDRSDFVADVAVAMTGG-----EGVDV-- 1161

```

```

Db      2817 RUTLAACDVADRASVAALLARLDAOGDPRAVAHVA--GGVVAQAALASTDLDCVQAVVA 2874
Qy      1162 -----INSLGELIDKSPNLIRSHGRVEIGKRCQYADNOLGRLPPLRNLSFSLVDLR 1214
Db      2875 KYRGALHTELLGDRLELDAFVLPAASGAGVWGSGQGAAYAGNAFLD-----ALAEVR 2926
Qy      1215 -----GMM-----LERPARVALLLELLGLLAAGV-----TP 1242
Db      2927 RAAGLTATSTIAMGAMAGGGLADHADADAEVELKRGGLVPMAPALIAALAGALDHGETA 2986
Qy      1243 PPIATLPARVADAFRMAQAQHLKVLVTLTGDPVEQIRIPT--HAGAGPSTGDRDLIDRL 1301
Db      2987 VVYASVDWARFAPAFASAPR-----PLLDLPKGAIAEAPAPAGADGASADRALVEAL 3042
Qy      1302 ASAAPARAAALEAFRTQVSQVLRTPETKVGAEAL-----PTRLGMSIMAVELENRRIE 1356
Db      3043 RPLSEADRAHVLVELVVAETAAVLGHAD-----ASALDPKGFPLDGLDILVAVELLRKLR 3098
Qy      1357 ASLKLTSTFLSTSPNIALAQNLDALATATLSLREVAENLRA 1401
Db      3099 ARTGVELPATVTFDHPSPHRLATYREALPLDAPLAPADATTARS 3143

```

## RESULT 10

```

OQ3NX8 PRELIMINARY; PRT; 5644 AA.
AC OQ3NX8;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE AmpHJ.
GN AMPHJ.
OC Streptomyces nodosus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=40318;
RN [1]
RA SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., O'Leary M.;
RT "The amphotericin biosynthetic gene cluster from Streptomyces
RL nodosus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases..
DR EMBL; AF57202; AAK73502.1; -.
DR InterPro: IPR001027; Ac transferase.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR004410; fadD.
DR InterPro: IPR000794; ketoacyl-synt.
DR InterPro: IPR006162; Pantane attach.
DR InterPro: IPR006163; Pp bind.
DR Pfam: PF00698; Acyl transfer. 3.
DR Pfam: PF00107; adh_zinc; 1.
DR Pfam: PF00109; ketoacyl-synt; 3.
DR Pfam: PF02801; ketoacyl-synt_C; 3.
DR Pfam: PF00550; pp-binding; 3.
DR TRGFAMs; TRGR00128; fadD; 3.
DR PROSITE; PSS0075; ACP DOMAIN; 3.
DR PROSITE; PSS00606; B KETOACYL SYNTHASE; 3.
DR PROSITE; PSS00012; PHOSPHOPANTETHEINE; 3.
DR Phosphotransferase; Transferease.
KW Phosphotransferase; Transferease.
SQ SEQUENCE 5644 AA; 585775 MW; CFEA35E3BA0D2658 CRC64;

```

Query Match 32.7%; Score 2357; DB 2; Length 5644;

Best Local Similarity 33.0%; Pred. No. 2.3e-106;

Matches 632; Conservative 216; Mismatches 509; Indels 558; Gaps 39;

```

Qy      11 EDPYIATGASGRLPGVYIDSGFTLLDGSRDTGVRVPAER--WDAAMPDPDDAPGKTP 69
Db      1797 DDVVIVYMGCRPGGSGSPEDLMDLLASGSDATGTPADREMESSLV--AGEAGSV- 1853
Qy      70 VTRASFSLDVACPDASFSGISPREALMDPAHRLLEVCWALENAAIAPSAIVGTGTG 129
Db      1854 -AOGFSLDIAGFDADFFGISPREALAMDPOQRITLLEVTWATERAGVDPTALRGSRGTG 1912

```

```

Qy      130 FIGIGPESEYEAALPOATASAEIDAHGSGCTMPSVGAGRISYALGRCPCVAVDTAVSSSL 189
Db      1913 FMVGAGQDSSLV--MGSRDDVAGHATAGLAVSVSGSLSTYGLGEPRLASTVDTCSSSL 1970
Qy      190 VAVHLACSLRSGECSTALAGVSLMSPSTLWLSKTRALARDGRCAFSADDFGFG 249
Db      1971 VSLHLAAQRLRGECSSMLVGGVTWTTTPANPAGFSRMGGLQADGRCAFSADAGTGS 2030
Qy      250 ECGAVVYLKRLSGAADDRILAVTSGAINHDGSSGLTVPNGSSOEVLKRLADAGC 309
Db      2031 EGAAYLVVERLSDARAGRIYAVVGSVANDGASNGITAPNGSQOQVIRIQALANGL 2090
Qy      310 AASSGVYEAHGTGTTGDPRIEQALNAYVGGRAVATPLIGSVKTNLGHREVASGITG 369
Db      2091 RGGDDVAVBAHGTGTPRGPIEAQALLATYSGDRPQCPILLIGSVKSNIGHTQAAAGVAG 2150
Qy      370 LKRVYLSLQHQIIPAHLAQALNPRI SWGDLRLTVTRATPWPMDNTPPRACVSSFGMSG 429
Db      2151 LVKMMAMRNGVLPRLTHITESTHYDMGLGAVQVLTETETAMPETGRVARAGVSSFGISG 2210
Qy      430 TNAHYVLEAPAACTPPAPERPABELV--LSATTAALDAQARLPHLETYSQCCG 486
Db      2211 TNAHYLEGADEPVPAVADRPVPGAVAMPVSASBGLDDQARLRASDALPALP-- 2267
Qy      487 DVAFSLATTSAMEHRLAVALAATSRGLRALDAAAGQTSPGAVERSIADSSRGKLAFLT 546
Db      2268 DTAITLATRADPEIR-----AVLLADG--TLTEVAGVAPRHS--AFIFS 2311
Qy      547 GGAQOTLGMGRGLYDVMSAFREAFDL CVRLFNOELDRPREVMAWBPASVDAALLDQTF 606
Db      2312 GGAQGLGMGRGLHAFPVFAEAFDSVTLLESELDTSVREVMW-----GTDEGALNATAF 2367
Qy      607 TOPALFTFERYALALMRSGVEPELVAGHSIGELVLAACVAVFSLEDVAFVLAARGRLMQ 666
Db      2368 TOPALFAVEVALYRLVESGVTPDFVAGHSVEIAAHVAGVFSLEDACRLVAARGRLMQ 2427
Qy      667 ALPAGAMVSIAPREADVAAVAAPHAASVIAVNAVDQVYAGAGQAPHAIAAAMAR 726
Db      2428 ALPAGAMVAVAETDEVEPLPUL--TEGVAVAANVQPSIVVSGEQATLVAEQALAAQ 2484
Qy      727 ARTKALVSHAFSPIMAELEAFGRVAVSVYRRSPYLVNSLSKACTDEV--SSPGW 785
Db      2485 RRTSRIRVSHAFSPIMAELEAFGRVAVSVYRRSPYLVNSLSKACTDEV--SSPGW 2544
Qy      786 VHAAREVAPFAGGVKALHAAGATFVEVGPKSTLLGLVAPCMDBARPALASSRAGRDP 845
Db      2545 VHAAREVAPFAGGVKALHAAGATFVEVGPKSTLLGLVAPCMDBARPALASSRAGRDP 2604
Qy      846 ATVLEALGLIMAVGLVSW--AGLFPSGGRVPLPTYPMQRERYW----- 888
Db      2605 VSVLSGVARLVYAGVDVDSAPLSGAGARIADVPTFAFCHERYWPAAPALDATGLGLA 2664
Qy      889 ----- 888
Db      2665 SADHPLIAGAAMSVAGSDELLITGSLSAATHPMLADHVVGMIFFPOTGLELAVRAADQA 2724
Qy      889 ----- 888
Db      2725 GCDRVEELMIAAPLVLPATGAVOVQISGAADBEGRBELFTPTPGEDPDAEMTOHATOR 2784
Qy      889 -----IDTKA-----DD----- 895
Db      2785 IGSGEVIDFDATWPPRAEALDIDGMFERAADGLEGVPRGLRAVMRQDDTYAAEV 2844
Qy      896 -----AANGDRRA----- 903
Db      2845 ALPESVEDADAFGLHALFDALHSTVFLSAEGDTRSLPFAWEGVSLHADGASTLRVRI 2904
Qy      904 -----PG----- 905
Db      2905 ASGGEVDVQIAAVDPGGQPVSVESLTLRAAGGDAAPRRDDDSNLLRVMTARTLGAAP 2964

```

```

Qy 906 -----
Db 2965 AARPATVAVLGEDPFGTLTAALAGDSEAVAGVHAPATIEELIARSGAVPDMVAVTVRGDAD 3024
Qy 906 AGHDEVEE-----
Db 3025 ACPDDAHLEHAHEVLALVQGMLEAPAPASSRLVVVTNNAAVADGERGAVDLAAAPVMGLVNS 3084
Qy 925 ARLDHP-----
Db 3085 AASENCRLLADVDOTADSLARPLLAGLFDABEEQAVVRECTVAVGKLARESTSLV 3144
Qy 944 --GDRPFLDEPGLVLDH--VLRVTERAPGLG--EVEIADVADAGLSENDVOLATLGMVD 999
Db 3145 PAGTFWRLGCRAGKSLDGLATLPPYEAATPLTGREVRIGIRAGNFRDVLNALGMYG 3204
Qy 1000 DLPGKNPPLLLGGEAGRIIVAVGSEVNLVVGQPIALSAQAFHVTTSALVLPQQ 1059
Db 3205 EAG-----LFGSEAGAVVSEGPVVTGLAPGDRVMGVFQGFGLVADERLLTRVD 3257
Qy 1060 ALASIEAAMPVAVLTAMVTLDRIARLQGERVLIHAATGVGLAIVQMAOHVGAEVHAT 1119
Db 3258 DMSWETASSVPLVLTATYALADLAGLRGEKLLIHAGAGVGMALIQLAHILGAEVAT 3317
Qy 1120 ACTPEKRAYLESIGVR--YVSDSRDPFVADVAMTGEVGVNLSLSEGLIDKSFNL 1177
Db 3318 A--SEGMVDVLRSGVADHDIASRTLDFTFAFTEVAGDKGLDVNLALAGEFVDSMRLL 3376
Qy 1178 BSHGRVLEIGKDCVADNOLGLRPFLRINSFSLVDLRGMLEBPAPVRLLEELGLITA 1237
Db 3377 GGGGRLENGKTDIRDSDAAS-----DGITYRFFDL--GMV--DPDHIQMLLDVLDLPER 3428
Qy 1238 GVTFPPPIATLPIARVADAFRSMAOQHLGKLVLTG--DPEVORITPTHAGA----- 1288
Db 3429 DVLSPPLVAVAMVDRREARFRRMSAKHIGKIVLTPRAMDEPGIVLVGGCGGLASALA 3488
Qy 1289 -----
Db 3489 RHLVAEHGIXHLLTSTRGPDAPGADLVQALAEABARVAAVDVADALAGLASAVP 3548
Qy 1330 IKVGAELFTRIGM--DSLMAVELNRRIEASLTKU-----ST 1365
Db 3549 AARPPLTAVVHTAGVLDGILASLTPRDLTVMRPKVDAAMHLDLTRDLAAFYLYST 3608
Qy 1366 TELSTSPNIALLA--ONLADALATALSERVAENL--RAGVONDFVSSG--ADOD 1415
Db 3609 SGVFGSGQANYAAGNTFLDALAAHQSLGIPATSLAMNAMBQSGMTGLSDSD 3663

```

RESULT 11  
Q9L4X2  
PRELIMINARY; PRT: 5435 AA.

Q9L4X2;  
01-OCT-2000 (Tremblrel. 15, Created)  
01-OCT-2000 (Tremblrel. 15, Last sequence update)  
01-OCT-2002 (Tremblrel. 22, Last annotation update)  
NYSJ  
GN  
OS Streptomyces noursei.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomyces; Streptomyces; Streptomyces.  
OX NCBI\_Taxid=1971;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 11455.  
RK MEDLINE=20334850; PubMed=10873841;  
RA Brattaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,  
RA Valla S., Zotchev S.B.;  
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in  
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and  
RT deduction of the biosynthetic pathway";  
RL Chem. Biol. 7:395-403(2000).  
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES

```

CC (SDR) FAMILY.
DR EMBL; AF263912; AAF71767.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001327; Ac_transferase.
DR InterPro; IPR002198; Adh_short.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; Fadd.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR006162; Pantane_attach.
DR InterPro; IPR006163; Pp_bind.
DR Pfam; PF00698; Acyl_transferf.
DR Pfam; PF00106; adh_short.
DR Pfam; PF00107; adh_zinc.
DR Pfam; PF00109; ketoacyl-synt.
DR Pfam; PF02801; ketoacyl-synt_c.
DR Pfam; PF00550; pp-binding.
DR TIGRfam; TIGR00128; fadd.
DR PROSITE; PS50075; ACP_DOMAIN.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE.
DR PROSITE; PS00012; PHOSPHOPANTHETINE.
DR Oxidoreductase; Phosphopantetheine; Transferase.
SQ SEQUENCE 5435 AA; 562659 MW; AA55465DF087A38C CRC64;

Query Match 32.7%; Score 2355.5; DB 2; Length 5435;
Best Local Similarity 34.2%; Pred. No. 2.6e-106;
Matches 655; Conservative 183; Mismatches 513; Indels 563; Gaps 47;

3 DPEIRAEADPIATVAGACRLPGVYIDLSGFVTLLEGSRDYGRPAER-WDAAMFDD 61
32 NEMLVAAAKEPIAVVAGMCRFPAGVDSPEALMEVATGTDVLSGPPDRGMLELRNSG 91
62 PDAPGKTPVTRASFVSDVACFDSPFGISPREALRMDPAHRLILLEVCMEALENNAIAPSA 121
92 TDARPTDVSQKRGFIDCIADPDRGFGISPREAVVMDPQRLLLTTAMEAVRAGIDATT 151
122 LVGTETGVPTIGIPSEYEALPQATASAEIDAAGLGTMPVSGAGRISYALGLRQCVAV 181
132 LRATRGAFITGNQDVAVYLIVRSLLDATTGV--GTGIAASAASGLSTYGLLEGPALTIV 209
182 DPAVSSSVAVHVLACQSLRSGECSTALAGCVSLMSPSTLVWLSTRALARPGRKAPSA 241
210 DTACSSSLVALHLLAVALQALRNGECGNALAGVAVMTPESLVFSSQGLARGRKAPAD 269
242 EADGFRGEGCAVVVLRKLSGARADDRILAVIRGSATINHDSAGSLTPVNSSQOEIVLK 301
270 AADGTGMEGACVLLERLSDQKRGHLYAVRRSAVNOGASNGFTAPNPSQQRVIR 329
330 QALANAGLATGIDDAVEAHGTGTPLGDPIEAQSILATV--GGDRAPVLLGSIKSNMGT 387
302 RALADAGCAASSVGVYEAHGTGTTLGDPIETIQALNAVYGLGRDVAATPLLIGSVKTNLGP 361
330 QALANAGLATGIDDAVEAHGTGTPLGDPIEAQSILATV--GGDRAPVLLGSIKSNMGT 387
362 EYASGITGLKLVLSLQGLPAHLHAQALNPRISW--GDRLTYTRARTPMPDNTPRR 419
388 QAAASVAGVYIKIMIMRHRGVLPRTLHVDRPSHYVMTGTSVEL-LTDAH-PWPERGRPRR 445
420 AGVSFGSGSTAAHVLEAP-----AATCPPAERPAELVLARSATSDAQAARLR 474
446 TGISFVGSGTAAHVLEAP-----AATCPPAERPAELVLARSATSDAQAARLR 504
475 DHELETPSQCLG---DVAFSLATTRSAMERHLAVATSREGL--RAALDAAAOCOTSPG 528
505 DHLDR-PGDRGPPTALDTAFSLATTRALAEHLAV--VTGTGTAGRDALTAVLHAGTARD 562
529 AVRSIADSSRGKLATLFTGCGAQTIGMGKGLYDVMSAREAVDLCVRLFNQSLDPLPRV 588
553 AHIEGHA-AGTRRCALFSGGQKORLGMREELHARPVPARALDVTAVDLDLGGTLEEV 621
589 MAAEPASVDAALDDOTAFOTPALFTFEYALALMRSMGVPELVAGHSIGELVACVAGV 648
622 IW---GTDDAPLANTGTGTQALPFAVEVALRLILSKGVAAPFVAGHSIGELVAAHVAGV 677
649 FSLSDAVFLVARGELMQALPAGAMVSIAPADVAALVAPHAASVSIAPVADQVVI 708

```



```

Db 678 FSLIEDACTVAAAGAGMOLPRGAMVAVEATEDEVSPILLTD---GVAIAINGPTSLV 734
Qy 709 AGAGGVVHIAAAMARGARTKALHSHASHPLMPLMEFGRVAVESVYRPPSYLV 768
Db 735 SGGETATLVAAALAOGRRTTRLRVSHASHPLMPMLAEFRAVAVESVYRPPSYLV 794
Qy 769 NLSGKACTDE-VSSPGVWVHAREVVRFADVKALHAAGAGTVEVGPSTLLGLVPACM 827
Db 795 NLTGAVADDTLGTADYWRHVRHVEAVRFADGIRALDAGVGLLELGPDDTLAALAOQSA 854
Qy 828 PDA-RPALLASSRAGDEPATVLEALGIMAVGLVSMGLFP-SGGRVRLPYTPWQR 884
Db 855 PDVAVSVVL---RKORDEEPAVAALALRLHTAGVPVDMTAFYAGTAGHRTDLPYAFQY 910
Qy 885 ERYWID-----TKADDAARD----- 900
Db 911 ERYWPKATYRPADATGTLTAADHPLLGAAMSVASDELLLTGTLSLATHPLADHVVGG 970
Qy 901 -----RRAPGAGHDEVEE----- 913
Db 971 MVEFPGTGLELAVRAADQVCDRVEHMLAEPILPATGTVMQIAVGAADDGGDRDLR 1030
Qy 914 -----G 914
Db 1031 FTRPGDDPDAAWAQHATGRITEGERVLTDTTTPPRDAEPYDIDGLYDRYANGIDYG 1090
Qy 915 GAARG-----GDRRS--- 924
Db 1091 PVERGLRAVWRDTEIYAEVALPBGTDADAFGLHPALFDVILHSTLPASADGDDSLRP 1150
Qy 925 -----ARLDH 929
Db 1151 FANNVSLHAAGADALRVRTITSCGPDAVEITAVDPQGRPVSVESLTLRAAGDAGADH 1210
Qy 930 -----PPESGRREKTEAAG- 944
Db 1211 RADAGSLFRMDWTPRTVHAAPATPATWALGTDPTGLTEALTAAGPDTVTGLRGDVALGE 1270
Qy 945 -----BRPF----- 948
Db 1271 LTAAGDRPVPDVAVPLRGATDHGPAGAHDLTRTVALLQEWLAEEFPAASRLILTRGA 1330
Qy 949 -----RLIEDP- 955
Db 1331 VADGERGPLDLAAPVWGLVRSASQSENRGRLLVDDDTAESAQLEPLRALLDADBPQA 1390
Qy 956 -----GVLDHL-VLRTERRAPGLG-EVE 977
Db 1391 VREGTVVRGLARLDSGRGLVPPPGTVMRLGSRAGSLDGLALLPPEARRPLTGHEVR 1450
Qy 978 IAVDAAGLSFNDVQLAGVPPDLRGKRNPLLGGSCAGRIYAVGCVGVAVGQVIA 1037
Db 1451 VGRAGLNFRLDNLNAGMTPGDAG-----LFGSAGAVVEVGEVGLAPGDWVG 1503
Qy 1038 LSAGAFATHTTSSAALVLRPOALSAIEAAMPAVATJAWALDRIRALOPGERVILTHA 1097
Db 1504 MLRGGFPEPLIGADARLLTLPVADMSWETGASVPLVFLTAAYALKELGGLAAGEVIAHAG 1563
Qy 1098 TGGVGLAAVMAOHVGAEVHATCTPEKRAVLESGLR--YVSDSRSDRVADVRAATGC 1155
Db 1564 AGGGAAMAIQIARVGAVEPATI-SEKMDVLRSLGVADHDIASRSLIDEFAAPAEVAGD 1622
Qy 1156 EGVADVNLNSLSELIDSFNLSHGRFVELGKDCVADNOLGLRPLRLNLSFVLVLRG 1215
Db 1623 RGLDVVNLNSLSEGVDSMRLLGDGRFLEWGTKDITPAADSVP-----DGLSHSFLP-G 1676
Qy 1216 MLEPAPVAVLLEELLGLIAAGVTPPPIATLPIARVADAFRSMAQOHLGKLVLTG- 1274
Db 1677 MV--DPHIQRMLLDVLELFDRGALALALPVRSMVVRAGAEFRMSLAQHLGKLVLTG 1734
Qy 1275 --DPEVOIRIPTHAG--AG-----DSTGDRDIL--DRLASAAPARA----- 1310
Db 1735 PLDPDGTVLTLTGCGGLARLRLVTEHGABRHLLLAGRGPRDAPGAALHAELTALGAE 1794

```

```

Qy 1311 ----AALEAFRTQVSOVLRTPEIK-----VGAELFTRLGMSLMAVELRNRIE 1356
Db 1795 VTVAAQDVADRTRALALATVPAEHPLTVAVHTAGVLDGDTLTALNPDLATV-LRPKD 1853
Qy 1357 AS-----LKLKSTFLSTS-----PNIALLA--ONLIDALAT--ALSL 1391
Db 1854 AAMHLLDLTRHLDLAFLVLYSSTAGVWGQGANVAAAGNTFLDALAAHRAALGL 1907

RESULT 12
ID Q914W3 PRELIMINARY; PRT; 11096 AA.
AC Q914W3;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE NYS.
GN NYS.
OS Streptomyces noursei.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brautaaset T., Sekurova O.N., Stetla H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zotchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway.";
RL Chem. Biol. 7:395-403(2000).
RL EMBL; AF263912; AAF71776.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FadD.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR006162; pantane_attach.
DR InterPro; IPR006163; pp_bind.
DR Pfam; PF00698; Acyl_transf. 6.
DR Pfam; PF00107; adh_zinc1. 1.
DR Pfam; PF00109; ketoacyl-synt. 6.
DR Pfam; PF02801; ketoacyl-synt_C. 6.
DR Pfam; PF00550; pp-binding. 6.
DR TIGRFAMs; TIGR00128; fadD. 6.
DR PROSITE; PS50075; ACP_DOMAIN. 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE. 6.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE. 5.
KW Phosphopantetheine; Transferase.
SQ
SEQUENCE 11096 AA; 1150415 MW; 776CAEAFCAE55IDD CRC64;

Query Match 32.3%; Score 2331.5; DB 2; Length 11096;
Best local Similarity 32.9%; Pred. No. 1e-104;
Matches 629; Conservative 182; Mismatches 519; Indels 583; Gaps 34;

Qy 10 AEDPIAVGASCRLLPGGVLDLGSFVTLGSRDVTGRVPAER-WDAAMFDPDPAPGKT 68
Db 3499 ADDPIVVGAGACFPFGGVTSPEDLMLRVDDGTDATITFPNKGMDLNDLXDPPEHFGTS 3558
Qy 69 PVTASFLSDVACFDASFGISPREALRMDPAHRLLEVCWEALENAAIAPSLVGTETG 128
Db 3559 YTRSGGFTHAGSEFDFAFGMSRRLATDSQGRLLLESWEAIEFAGIDPLTRGSATG 3618
Qy 129 VFTIGDSVEYEAALPQATASAEIDAAGLGCTMPSVAGAGRIAYLGLRGVAVDTAYSSS 188
Db 3619 VFAGWYSDGYSIL-----GGEKEFGQGGGSASVSGRVSYALGEGPAVTVDTACSSS 3674
Qy 189 LVAVHLACSLRSBGESTLAGVSLMLSPSTLWLSKTRALARDGRCKAFSAEADQFGR 248
Db 3675 LVALHMAAQALRLRAGCSLALAGGVTVWSTPSTFVFSRQGLAPDGRSAFAEADQGVW 3734

```



DR Pfam: PF00106; adh\_short; 3.  
 DR Pfam: PF00109; ketoacyl-synt\_C; 6.  
 DR Pfam: PF02801; ketoacyl-synt\_C; 6.  
 DR Pfam: PF00550; pp-binding; 6.  
 DR TIGRfam: TIGR00128; fadD; 6.  
 DR PROSITE: PS50075; ACP DOMAIN; 6.  
 DR PROSITE: PS00606; B KETOACYL\_SYNTHASE; 6.  
 DR PROSITE: PS00037; MVA 1; 1.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 6.  
 DR Oxidoreductase; Phosphopantetheine; transferase.  
 KW OXIDOREDUCTASE; 9507 AA; 994228 MW; 57BD80CBAD37E89F CRC64;  
 SQ

Query Match 32.3%; Score 2330; DB 2; Length 9507;  
 Best Local Similarity 37.6%; Pred. No. 9,9e-105;

Matches 624; Conservative 179; Mismatches 544; Indels 312; Gaps 40;

QY 9 AADPIAIVAGSCLPLPGVITLDSFWTLLEGSRDVTGVPAER-WDAANFPDPPDAPGK 67  
 DB AADPIAIVAGSCLPLPGVITLDSFWTLLEGSRDVTGVPAER-WDAANFPDPPDAPGK 67  
 QY 68 TPTTRASFSDVACFPDASFSGISPRELMDPAHRLLEVCWEALEMAALAPSAVGTET 127  
 DB TPTTRASFSDVACFPDASFSGISPRELMDPAHRLLEVCWEALEMAALAPSAVGTET 127  
 QY 4916 TVSTRGGFLHDAADFDASFGISPRELMDPOORLLLETSMEMERAGIDPATLRGSSST 4975  
 DB TVSTRGGFLHDAADFDASFGISPRELMDPOORLLLETSMEMERAGIDPATLRGSSST 4975  
 QY 128 GVFIIGIPSEYEA--ALPQATASAEIDAAGLGTMPSVGAGRI SYALGLRGPVADVTAY 185  
 DB GVFIIGIPSEYEA--ALPQATASAEIDAAGLGTMPSVGAGRI SYALGLRGPVADVTAY 185  
 QY 4976 GTITGASDYDYRAFGAGPDGA-----EGHILITGIVSLSGRLSYTTGFGFPVSDITAC 5030  
 DB GTITGASDYDYRAFGAGPDGA-----EGHILITGIVSLSGRLSYTTGFGFPVSDITAC 5030  
 QY 186 SSSLVAVHLACQSLIRNGESSLALAGVSLMSPSTVWLKSTRALARDGRCKAFSAEADG 245  
 DB SSSLVAVHLACQSLIRNGESSLALAGVSLMSPSTVWLKSTRALARDGRCKAFSAEADG 245  
 QY 5031 SSSLVAVHLACQSLIRNGESSLALAGVSLMSPSTVWLKSTRALARDGRCKAFSAEADG 5090  
 DB SSSLVAVHLACQSLIRNGESSLALAGVSLMSPSTVWLKSTRALARDGRCKAFSAEADG 5090  
 QY 246 FPGEGEAGVAVVRLSARADGRIILAVIRGSAINHGAASSGLTPVPGSOEIVLAKALA 305  
 DB FPGEGEAGVAVVRLSARADGRIILAVIRGSAINHGAASSGLTPVPGSOEIVLAKALA 305  
 QY 5091 MSIAEGVALVRLSDARRNGHQVLAVIRGSAINHGAASSGLTPVPGSOEIVLAKALA 5150  
 DB MSIAEGVALVRLSDARRNGHQVLAVIRGSAINHGAASSGLTPVPGSOEIVLAKALA 5150  
 QY 306 DAGCAASSVGVYEAHGTGTLGDPRIEIQALNAVYGLRDVATPPLLIGSVTNLGHPEYAS 365  
 DB DAGCAASSVGVYEAHGTGTLGDPRIEIQALNAVYGLRDVATPPLLIGSVTNLGHPEYAS 365  
 QY 5151 NAEVDPDGIIVIDHGHGTALGDPRIEIQALNAVYGLRDVATPPLLIGSVTNLGHPEYAS 5210  
 DB NAEVDPDGIIVIDHGHGTALGDPRIEIQALNAVYGLRDVATPPLLIGSVTNLGHPEYAS 5210  
 QY 366 GITGLKLVLSLHOGQIPAHLLAOALNPRISWQDLTLTVTRAPWDMNTPRAGVVSF 425  
 DB GITGLKLVLSLHOGQIPAHLLAOALNPRISWQDLTLTVTRAPWDMNTPRAGVVSF 425  
 QY 5211 GVASVIMVMAMRHAAPVRLHIDRPSHVMWSGAIQLTEPLPMPATGCHPRAGVVSF 5270  
 DB GVASVIMVMAMRHAAPVRLHIDRPSHVMWSGAIQLTEPLPMPATGCHPRAGVVSF 5270  
 QY 426 GMSGTAAHVLEAP---AATCTPRAPERAE---LLVLSARASLMDAOARLRLHLE 478  
 DB GMSGTAAHVLEAP---AATCTPRAPERAE---LLVLSARASLMDAOARLRLHLE 478  
 QY 5271 GLSGTNVHTLLEQAPEDRTEDDTAPARADPTVPVLTLSGRTEAGLRQAGRLTLHLA 5330  
 DB GLSGTNVHTLLEQAPEDRTEDDTAPARADPTVPVLTLSGRTEAGLRQAGRLTLHLA 5330  
 QY 479 TYSQCLGDVAFSIATRSAMEHRLAVATSRREGIARAALDAAAGQTS PAVRSIADSSR 538  
 DB TYSQCLGDVAFSIATRSAMEHRLAVATSRREGIARAALDAAAGQTS PAVRSIADSSR 538  
 QY 5331 DHEBELPLTDLAFSLATSRALHRAAVVTTDDPTLTALTRALNDATPDGGLTGRPD--R 5388  
 DB DHEBELPLTDLAFSLATSRALHRAAVVTTDDPTLTALTRALNDATPDGGLTGRPD--R 5388  
 QY 539 GKLAFLFTGGCAQTLGKRGVLYVWSAFREAFDL CVRLFNOELDRPLREVMMAEPASVDA 558  
 DB GKLAFLFTGGCAQTLGKRGVLYVWSAFREAFDL CVRLFNOELDRPLREVMMAEPASVDA 558  
 QY 5389 GRILAFLFTGGCAQTLGKRGVLYVWSAFREAFDL CVRLFNOELDRPLREVMMAEPASVDA 5448  
 DB GRILAFLFTGGCAQTLGKRGVLYVWSAFREAFDL CVRLFNOELDRPLREVMMAEPASVDA 5448  
 QY 599 ALLDDQTAFTQPALFTFEYALALALMRSGVBEVLVAGSISGEIVAACVAFSLDEAVFLV 658  
 DB ALLDDQTAFTQPALFTFEYALALALMRSGVBEVLVAGSISGEIVAACVAFSLDEAVFLV 658  
 QY 5449 ELDDDTGYTQPALFTFEYALALALMRSGVBEVLVAGSISGEIVAACVAFSLDEAVFLV 5508  
 DB ELDDDTGYTQPALFTFEYALALALMRSGVBEVLVAGSISGEIVAACVAFSLDEAVFLV 5508  
 QY 659 AARGRLMALPAGGAMVSIAREADVAAVAAPHAASVIAAVNAPDOVTLAGAGQPVHAI 718  
 DB AARGRLMALPAGGAMVSIAREADVAAVAAPHAASVIAAVNAPDOVTLAGAGQPVHAI 718  
 QY 5509 AARGRLMALPAGGAMVSIAREADVAAVAAPHAASVIAAVNAPDOVTLAGAGQPVHAI 5568  
 DB AARGRLMALPAGGAMVSIAREADVAAVAAPHAASVIAAVNAPDOVTLAGAGQPVHAI 5568  
 QY 719 AAAAAGARTKALHVSASHFSPMLAPLEAFGRVAESVSRISIVLSNLGK--ACTD 777  
 DB AAAAAGARTKALHVSASHFSPMLAPLEAFGRVAESVSRISIVLSNLGK--ACTD 777  
 QY 5569 AAHFHELGRTKQLRVSHAFSPHMDMLDFTRAAGITFAHABELPLVSTLTGELLAD 5628  
 DB AAHFHELGRTKQLRVSHAFSPHMDMLDFTRAAGITFAHABELPLVSTLTGELLAD 5628  
 QY 778 EVSGPGVWVHARVAVRFADGVKALHAAGACTFVEVGPSTKLGLVAPCAHPDARPA--LLA 836  
 DB EVSGPGVWVHARVAVRFADGVKALHAAGACTFVEVGPSTKLGLVAPCAHPDARPA--LLA 836  
 QY 5629 DLCSAPFWTRQVTRGVFAADAVRLTGHVTTFFELPDVLTSGAVESSTGEQTPAVAP 5688  
 DB DLCSAPFWTRQVTRGVFAADAVRLTGHVTTFFELPDVLTSGAVESSTGEQTPAVAP 5688  
 QY 837 SSRAGRDEPATVLEALGLMAVGLVSMAGLFPSSG--RAVPLPTYPWQRRRYWIDTAD 895  
 DB SSRAGRDEPATVLEALGLMAVGLVSMAGLFPSSG--RAVPLPTYPWQRRRYWIDTAD 895

DB 5689 ALRRORPEAPALTTALALHLGVRVMDAVFAGGARIDLPYFQGRERYWPEAAAGA 5748  
 QY 896 AA-----RGDRRAPAGAGHD-----EVEGGAVRG 920  
 DB 5749 AAPAPADAAAEFMSAVERADLPISGSSLDLDDDTLTAVVPAVLSWRKRSRSTVDGM 5808  
 QY 921 DRR-----SARLDHP-----PESGREKEVEA-----GDRPFRLI 952  
 DB 5809 KIRTTKPLTGNITAGHPAGTGLVLAAPAGDDAAMDVAAGALGADAVRVEVTAQOQLA 5868  
 QY 953 -----DEPGVL-----DHLVLRVTERRAPGLGEVE----- 977  
 DB 5869 ERLSELAAHGEFAGVLSLATAEGEADADATTEGLLTATATLQALDAGIDAPLMCV 5928  
 QY 978 ----IYVDAAGSPNDVOLA---LGVN-----PDLLPKGNPRLILGECACRIYAVGSG 1025  
 DB 5929 TRTAVAVDAAEHAPAPQAQAAVWGLGRVAALEHPQWVGSLVLDLPDELIDGATLRLAA--- 5984  
 QY 1026 VNGLVVGGPVIALSAGAFATHTTSAAVLPRPOALSAIEAAMPVAVYLTAWYALDRIRAR 1085  
 DB 5985 -----VLADSGEDQLAARTTATFV-----RLAHHRAAPAPA-----AT 6020  
 QY 1086 LQGEREVLJHAATGGVGLAAVQAGVGAENHA-----TAGTPEKRAYLESIGVRY 1136  
 DB 6021 PRPTGTVLVTGGTALGHHVAVRLABAGAE--HLLVSRGRTDAPGAGELAAEIAELGARV 6079  
 QY 1137 ---VSDRSRDFRVDYRA-----WTCGEVDVVLNLSGSE----- 1168  
 DB 6080 TYAACCTDADALAAVLAALPDEHPLTAVFHTAGTVDDGLTDLTLPQEPASVLRKYAT 6139  
 QY 1169 -----LIDKSFNLRSH-----G 1181  
 DB 6140 RNILHETTRRODLSAPVLFSSVAGTGLAPGQGVYAAAGNAFLDPAFAHRRARHGLPATSLANG 6199  
 QY 1182 RYVELGKRDYADNQLRPPRLNLSFSLVDLRGMMLERPARVALLLEGLIAAG--- 1238  
 DB 6200 PMAETG---MAADGTGIDRVRGCGFTMPPR---LALTARALAEHDATATLTADIDW 6252  
 QY 1239 ---VFTPPPLATLPIARVADAFRMAQAQHLGKVLTLGSPDEVIRIPTAGAGPSTG 1293  
 DB 6253 QRYAEVFT---ATRSLSFVAD---LPELQOVTTPRAGTPEAALREBA----- 6292  
 QY 1294 DRDLRLASAPARAALAEALFRTQVQVL--RTPEIKVGAELFTRLGMDSLMAVEIR 1352  
 DB 6293 ---LRQLAGLSAPAPRFLVDLVRQVAAVIGHSTSAIGABRARSIDGFPSTLTVEIR 6349  
 QY 1353 NRIEASLKULSTFLSTSPNIALAQNLLDALATLS-----LERVAAE----- 1397  
 DB 6350 NTLTATTGKLPLATLVYDPTPALADFLIAELIAGLPESGAPAFVGRAADDPPIVIGM 6409  
 QY 1398 NLR--AGVON-----DFVSSG-----ADDMETIAL 1421  
 DB 6410 NCRFPGGRSPEDLMQLSSGEDAISGFPRADKMDIDL 6448

## RESULT 14

Q93NM6 PRELIMINARY; PRT; 10917 AA.  
 AC Q93NM6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE AmpHC.  
 GN AmpHC.  
 OS Streptomyces nodosus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomyces; Streptomyces; Streptomyces.  
 OX NCBI\_TaxID=40318;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Caffrey P., Lynch S.V., Flood E.M., Finnian S.M., O'Leary M.,  
 "The amphotericin biosynthetic gene cluster from Streptomyces



QY 1292 TGDRLDLRLASAPARAALAEFLRTQVSOYLRTPEIKVGAALFTRLGMSLMAVEL 1351  
 DB 5204 TGTGTGLGGL-----VARHLYMERGVRRLL-----LTSRGLDLAGARREL 5243  
 QY 1352 RNRIEASLKLKLTSTFLSTSPNIALLAQNLLDALATALSLERVAEN-LR-----AGVON 1405  
 DB 5244 VAELE-----NLGAEVSVAACDVADBDVAVDTLLAGIPAEHPLBAVHTAGVLD 5291  
 QY 1406 DFV 1408  
 DB 5292 DGV 5294

RESULT 15  
 ID 053490 PRELIMINARY; PRT; 4151 AA.  
 AC 053490;  
 DT 01-JUN-1998 (TREMREL.06, Created)  
 DT 01-JUN-1998 (TREMREL.06, Last sequence update)  
 DT 01-MAR-2003 (TREMREL.23, Last annotation update)  
 DE Polyketide synthase.  
 GN PKS12 OR RV2048C OR MT2108 OR MTV018.35C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1773;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,  
 RA Hornsby S., Jagels K., Krogan A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulten J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J.F., Deboy R., Dodson R., Gwim M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL021899; CAI17262.1; -  
 DR EMBL; AB007061; AAK46387.1; -  
 DR HSSP; P25715; IMLA.  
 DR TIGR; MT2108; -  
 DR Tuberculist; RV2048C; -  
 DR InterPro; IPR001227; Ac\_transferase.  
 DR InterPro; IPR002085; Adh\_zn\_family.  
 DR InterPro; IPR00794; ketoacyl-synt.  
 DR InterPro; IPR006162; Pantane\_attach.  
 DR InterPro; IPR006163; Pp\_bind.  
 DR Pfam; PF00698; Acyl\_transf. 2.  
 DR Pfam; PF00107; adh\_zinc. 2.  
 DR Pfam; PF00109; ketoacyl-synt. 2.  
 DR Pfam; PF02801; ketoacyl-synt. C; 2.  
 DR Pfam; PF00550; pp-binding. 2.  
 DR PROSITE; PS00705; ACP\_DOMAIN; 2.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 2.  
 DR PROSITE; PS00012; PHOSPHOPANTHEINE; 2.  
 DR Phosphopantetheine; transferase; Complete proteome.  
 KW Phosphopantetheine; transferase; Complete proteome.  
 FT CONFLICT 2147 2147 H -> Q (IN REF. 2).  
 FT CONFLICT 2260 2261 DI -> TV (IN REF. 2).

FT CONFLICT 2268 2268 W -> H (IN REF. 2).  
 FT CONFLICT 2272 2272 S -> A (IN REF. 2).  
 FT CONFLICT 2279 2280 AF -> PY (IN REF. 2).  
 FT CONFLICT 2282 2283 AA -> GR (IN REF. 2).  
 FT CONFLICT 2287 2287 T -> V (IN REF. 2).  
 FT CONFLICT 2289 2289 F -> W (IN REF. 2).  
 FT CONFLICT 3004 3004 S -> L (IN REF. 2).  
 FT CONFLICT 3649 3649 P -> A (IN REF. 2).  
 SQ SEQUENCE 4151 AA; 431577 MW; 536F64F111D7AD CRC64;

Query Match 32.2%; Score 2321; DB 16; Length 4151;  
 Best Local Similarity 33.2%; Pred. No. 8.7e-105;  
 Matches 613; Conservative 203; Mismatches 535; Indels 498; Gaps 31;

QY 6 IERAAEDPRAIVGASGRLPGGVLDLGGFTLLGSGRDTGVRPAER-WDAAMFDDPPA 64  
 DB 2051 VRTTSEDPPIAIVGMACRYPGVNSPDDMDMLIQGDVLSERFPDGLGYNDDPA 2110  
 QY 65 PKTPTTRASFSDVACFDASFFGISPREALRMDPAHRLLEVCWALENAAIAPSAVG 124  
 DB 2111 AGACTRTGTFDVGDFDPAFFGVGSEALAMDPOHRLLELSWEALERAGIDPTGLRG 2170  
 QY 125 TETGVFTIGPSEYEALPQATPAEIDAHGGLGTPSGAGRISTALGRPCAVDTA 184  
 DB 2171 SATGVFAGVWTOGY-----GMFAAEPEVEGFRULGQLSVAAGVAVYGLGPAVSVDTA 2225  
 QY 185 YSSSLVAVYLAOCSTLRSGECSTALAGVSLMSPSTLWLSKTRALARDGRCKAFSAED 244  
 DB 2226 CSSSLVALHVAAGSLSGECDLALAGVTVNATPDIFVFESKRGLSPGRCAPKAAD 2285  
 QY 245 GFGRGCAVAVVVKRLSGARADGDRILAVIRGSAIINHDSAGSLTVPNGSSOEIVLKRAL 304  
 DB 2286 GTFSGSGGMLVYQRSDARLGHPLAVAVGSVAVNQDASNGLTAPNGPQORVVRAL 2345  
 QY 305 ADHGAASSVGYEAHGTGTTCADPIETALNAVYGLG-DAVPLIGSVKTNLGHPEY 363  
 DB 2346 ANAGLSAAEVDVVEGHTGTTCADPIEADALLATYQDDEBPDEPLMLSVXNMGHTQA 2405  
 QY 364 ASGITLLVYLSLQSGQPAHLHAQALNPRIISWGLRLTVYRATPMDWMTPRAGVS 423  
 DB 2406 AAGVAVYIKVYLMRHELLPATLHVDPSPHDMAGAVELTAPVYRAGATRRAGVS 2465  
 QY 424 SFGMSGTNAHVLEAPPACTPPAPERPAELVLSARTASALDAQARLRLHLETPSQ 483  
 DB 2466 SFGISGTNHVIEIENPVPRRAGWAGVVPVNVSAKSESALRGQAARLAAYVRDDGL 2525  
 QY 484 CLQDVAFSLATRSAMEHRLAVALISREGLRALDAAGQCTSPAVRSIADSSRKLAF 543  
 DB 2526 DVADVGMSTLA-GRSVEHRAVAVVGSDRDLGLDLADDOGLGSVVRGTA-TAAKTYF 2583  
 QY 544 LFTGGAQNTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMMAEPASVDALLDQ 603  
 DB 2584 VFPQGQSQMLGKMGSLHAGYVPVFAEAFNTVYVGEIDHLRLPLREVMMGH-----DENLNS 2639  
 QY 604 TAFTOPALFFFEYALALMRSMGVPEBELVAGHSIGELVAAVCYAVGFSLEDAVEVLVARGR 663  
 DB 2640 TEFAQPALFAVEVALFRLLGSMGVDRDPVWGHISIGLSAHHVGVSLLENVAVLVAARGR 2699  
 QY 664 LMQALPAGAMNYSIEAPADVAAVAAPHAASVIAAVNAPDQVVIAGAQPVYAIATAAA 723  
 DB 2700 LMQALPAGAMVAVQAEEVERPL--SAEVDIAVNGVASLIVISGAQNAVAVAVDQLR 2756  
 QY 724 ARGARTKALVSHAFFSPLMAPLEAFGRVAESVSRPSIVVSNLSGKACTDEVSPPG 783  
 DB 2757 ADGRRVHQALVSHAFFSPLMDPVIDERPAVAGIAGRPITIGVISVNTGQLAGDDGSA 2816  
 QY 784 YVYHAREVVRFPADGVKALAAAGCTFVEVGPSTLLGLVPACMPARPALLASSBARD 843  
 DB 2817 YMRHRHROAVRFADSVRFQAAGSRFLVGPSSGLVASIEESLPDVAATVMTGALRLDRP 2876  
 QY 844 EPATVLEALGGLMAVGLVSWAGLPPSGGRVPLPTYPWOREKWT--DTKADDA----- 897  
 DB 2877 EPATLITNAVAGGVTCMDLDMRAVV-GEAQFVELPTYAFORRRFVLSGDDVADDAAGLGL 2935

```
QY 898 -----RGDRRAPAGH-----DE 910
Db 2936 AASEHLLGAVIDLPAAGGVLTGRSPSVQGLADHSAVAGVTIFPGAGFVELAIRAGDE 2995
QY 911 V-----EFG-----GAV 917
Db 2996 VCGGVDESTLAFLVLPASGSVAQVVVNGPDESGVSVYSGRDVGTVGLHAEGAL 3055
QY 918 RGS----- 920
Db 3056 RAGSAEPTLDLAMPAPGAVPEVEADGYOQLAERGVCYGPAPRGLTAMWRGDEVEFAEVA 3115
QY 921 ----- 920
Db 3116 LPADAGSVTGCVHPVLLDAALHAHVLSAESAEERGQSVLPFSGVSLHAAGASAVR 3175
QY 921 -----DR----- 922
Db 3176 ARIAPVPSAVSIELADGLPVLVSASMLARPVTQDLRAAVSSSGPRLFEVTSPOP 3235
QY 923 ----- 922
Db 3236 SAAVEPLPVCAAGTDESAAVFESVPLAGDVVAGVAAATSSVLDVLSMLTRDGAGLV 3295
QY 923 -----RSARLDHP-----P 931
Db 3296 VMTRGVALPGEDVTDLAGAAVWGLVRSAGTEHPRIVLVDSAPLDDSALAAVTTTGP 3355
QY 932 PESGRREKVEA-----GRRPRLIDEPGVLDHLVLR-VTERRAP-GLG 974
Db 3356 QVLMRGEVYRTARVHSGRAVGLLVPSPDRPWLAMSTAGTFENLRLELIPDAAPLPG 3415
QY 975 EVELIADAGLSFNDVOLALGMWPDPLPKRNPRLLLGECACRIYAVGEGVNGLVVGP 1034
Db 3416 QVRVAASAIANFRDWMIALGLYPD-----PDAVWGVEACGVLETSLKSGFAVGD 3468
QY 1035 VIALSAGAFATHVTTSAALVLRPPALSAIEAAMPAYLTAMAYALDRIARLOPGERVLI 1094
Db 3469 VMGLPREGTGYASTQRLLVKVPAGMSTHAAATTSVPATAHYALVDLAAARSQGVLI 3528
QY 1095 HAATGGVGLAAVQMAQHVGAEVHATGTEPKRAYLESIGV-RYVSDSRSDRFVADVRAM 1152
Db 3529 HAGTGGVGAAYQVLAHHLGLEVPATA-SKGKMDTLRAMGFDDDHISDSRSLFEDEKRAA 3587
QY 1153 TGGEQVDVYVLSIGELIDKSFNLSRSHGFVELEGKDCYADNQLGR-PFLRNLSFSLV 1211
Db 3588 TGRGRFDVYLDISLAGEFVDASLRVAPGVFLEMGTDIRDPGVIAQYFGVRYRAFDF 3647
QY 1212 DLRGWMLERRPARVRLLEELGLIAGVFTPPPIATFLPIARVADAFRSMQAQHLGKLV- 1270
Db 3648 E-----PGPDRIAQILAEIATLFGDGLRPLFVTTFDVRCAPALRLYLSQARHTGKVM 3701
QY 1271 -----LTTGDPEVOIRIPTHAG-----GP-STGDRDLRLLAGA 1304
Db 3702 LMPGSWAAGTVLITGGTGMAGSAVARHGVARVLYSRGPDAPGAELVAELAA 3761
QY 1305 APAARAAALEAFLRTQVSOVLKTEPK-----VGA--EALFTRLGMDSLMAVELRN 1353
Db 3762 GAQOVVACDADRALAKAVIADIPVQHPLSGVITHAGALDDAVVMSLTPDRVDV-IRS 3820
QY 1354 RIEASLKLSTFTLSTSPNIAL-----LAQNLDPALA 1386
Db 3821 KYDAAMHHLTRDLVDVSAFVMSMAGLVGSSGQANYAANGFLDALA 3869
```

Search completed: October 2, 2003, 17:31:44  
Job time : 167.882 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 26.2576 Seconds  
(without alignments)  
2544.975 Million cell updates/sec

Title: US-10-014-717-2  
Perfect score: 7210  
Sequence: 1 VADRIERAEADPIAIVGAS.....GVQNDPVSSGADQWEIHAL 1421

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries.

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2107.5	29.2	3567	1 ERY2_SACER	003132 saccharopol
2	1925.5	26.7	3491	1 ERY1_SACER	003131 saccharopol
3	1889	26.2	3519	1 OL56_STRAT	007017 streptomyc
4	1875	26.0	3172	1 ERY3_SACER	003133 saccharopol
5	1852.5	25.7	2110	1 MCAS_MYCBO	002251 mycobacteri
6	1592.5	22.1	1538	1 PRSB_MYCTU	010578 mycobacteri
7	1583	22.0	1876	1 PSA_MYCTU	010577 mycobacteri
8	1180	16.4	1986	1 WA_EMENT	003149 emericella
9	1177	17.4	1774	1 MSAS_PENPA	P22367 penicillium
10	1173.5	16.3	2109	1 PKSI_ASPPA	012053 aspergillus
11	1100	15.3	2505	1 PAS_FAP	P12785 ratius novy
12	1083	15.0	2181	1 STCA_EMENT	012397 emericella
13	1039	14.4	2504	1 PAS_HUMAN	P49327 homo sapien
14	1018	14.1	4447	1 PKSR_BACSU	P40803 bacillus su
15	997.5	13.8	2511	1 PAS_CHICK	P12776 gallus gall
16	940	13.0	4427	1 PKSI_BACSU	005470 bacillus su
17	882.5	12.2	4273	1 PKSM_BACSU	P40872 bacillus su
18	612.5	8.5	352	1 PAS_ANGAN	P36189 anser anser
19	392.5	5.4	317	1 FAD_BACSU	P71019 bacillus su
20	363	5.0	401	1 NOE_RHIME	P06331 rhizobium m
21	354.5	4.9	402	1 NOE_RHIME	P06330 rhizobium m
22	344	4.8	308	1 FAD_SALTY	O85140 salmoneila
23	336	4.7	403	1 NOE_RHIVV	P04683 rhizobium l
24	332	4.6	308	1 FAD_ECOLI	P25715 escherichia
25	329.5	4.6	467	1 KASA_STRCO	Q02059 streptomyc
26	328	4.5	312	1 FAD_HAEIN	P43712 haemophilus
27	325.5	4.5	402	1 NOE_RHIS3	P72331 rhizobium s
28	324.5	4.5	401	1 NOE_RHILT	P04684 rhizobium l
29	321.5	4.5	426	1 KASI_STRGA	P16538 streptomyc
30	316.5	4.4	300	1 VARI_HUMAN	Q99536 homo sapien
31	315	4.4	410	1 FAD_ECOLI	P39435 escherichia
32	308	4.3	414	1 FAD_VIBHA	P55338 vibrio harv
33	304.5	4.2	293	1 FAD_SYNY3	P73242 synecocyst

34	303.5	4.2	413	1 FAD_VIBCH	Q9kq9 vibrio chol
35	303	4.2	838	1 PAS_MOUSE	P19096 mus musculu
36	302	4.2	420	1 KASI_STRCM	P41175 streptomyc
37	298.5	4.1	327	1 OOR_ECOLI	P28304 escherichia
38	298	4.1	421	1 KASI_STRVN	P16540 streptomyc
39	295.5	4.1	327	1 OOR_SALTY	P40783 salmoneila
40	293	4.1	421	1 FAD_RHIME	P56902 rhizobium m
41	289.5	4.0	422	1 KASI_STRHA	Q05356 streptomyc
42	285.5	4.0	416	1 FAD_SYNY3	P73283 synecocyst
43	284.5	3.9	423	1 KASI_STRCO	P23155 streptomyc
44	282	3.9	379	1 VARI_TORCA	P19333 torpedo cal
45	278	3.9	325	1 OOR_PSEAE	P43903 pseudomonas

## ALIGNMENTS

RESULT 1  
ID ERY2\_SACER STANDARD; PRT; 3567 AA.  
AC 003132; 054096;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-deoxyerythronolide B synthase II) (DBS 2).  
GN BRVA.  
OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Pseudonocardiales; Pseudonocardaceae; Saccharopolyspora.  
OX NCBI\_TaxId=1836;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91220065; PubMed=2024119;  
RA Donadio S., Staver W.J., McAlpine J.B., Swanson S.J., Katz L.;  
RT "Modular organization of genes required for complex polyketide biosynthesis.";  
RL Science 252:675-679(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL 2338;  
RX MEDLINE=92155230; PubMed=1740151;  
RA Bevil D.J., Cortes J., Haydock S.F., Leadley P.F.;  
RT "6-deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme.";  
RL Eur. J. Biochem. 204:39-49(1992).  
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-deoxyerythronolide B.  
CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHEINES.  
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN BIOSYNTHESIS.  
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6, RESPECTIVELY.  
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT), BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR), DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION OF THE FULL-LENGTH CHAIN.  
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (PAS).  
CC -1- SIMILARITY: Contains 2 acyl carrier domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M63677; AAA26494.1; -

DR EMBL; X62569; CAA44448.1; -

DR PIR; S23070; S23070.

DR InterPro; IPR001227; AC transferase.

DR InterPro; IPR002085; Adh zn family.

DR InterPro; IPR000794; Ketocacyl-synt.

DR InterPro; IPR006163; Pp\_bind.

DR InterPro; IPR006162; Ppanthe\_attach.

DR Pfam; PF00698; Acyl\_transf; 2.

DR Pfam; PF00107; ADH zinc N; 1.

DR Pfam; PF00109; ketocacyl-synt; 2.

DR Pfam; PF02801; ketocacyl-synt; 2.

DR Pfam; PF00550; pp-binding; 2.

DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.

DR PROSITE; PS00606; B KETOACYL SYNTHASE; 2.

DR PROSITE; PS50075; ACP DOMAIN; 2.

KM transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;

KM Phosphopantetheine; Multifunctional enzyme.

FT DOMAIN 1 1484 MODULE 3.

FT DOMAIN 1485 3567 MODULE 4.

FT DOMAIN 27 488 BETA-KETOACYL SYNTHASE 1.

FT DOMAIN 559 884 ACYLTRANSFERASE (AT) 1.

FT DOMAIN 1130 1301 BETA-KETOACYL REDUCTASE 1 (POSSIBLY NON-FUNCTIONAL).

FT DOMAIN 1397 1467 ACYL CARRIER (ACP) 1.

FT DOMAIN 1485 1943 BETA-KETOACYL SYNTHASE 2.

FT DOMAIN 2013 2336 ACYLTRANSFERASE (AT) 2.

FT DOMAIN 2383 3066 DEHYDRATASE/ENOYLREDUCTASE (DH/ER).

FT DOMAIN 3139 3322 BETA-KETOACYL REDUCTASE 2.

FT DOMAIN 3415 3485 ACYL CARRIER (ACP) 2.

FT ACT\_SITE 202 202 THIOESTER BOND.

FT ACT\_SITE 651 651 ACYL-ESTER INTERMEDIATE.

FT BINDING 1430 1430 PHOSPHOPANTHETHEINE (BY SIMILARITY).

FT ACT\_SITE 1661 1661 THIOESTER BOND.

FT ACT\_SITE 2105 2105 ACYL-ESTER INTERMEDIATE.

FT NP\_BIND 2961 2978 NADP (ER).

FT NP\_BIND 3142 3157 PHOSPHOPANTHETHEINE (BY SIMILARITY).

FT BINDING 3448 3448 NADP (KR).

FT BINDING 438 438 R -> A (IN REF. 2).

FT CONFLICT 480 480 T -> S (IN REF. 2).

FT CONFLICT 1241 1241 L -> F (IN REF. 2).

FT CONFLICT 2664 2664 G -> V (IN REF. 2).

SO SEQUENCE 3567 AA; 374413 MM; EE6284F4738AA0C0 CRC64;

Query Match 29.2%; Score 2107.5; DB 1; Length 3567;

Best Local Similarity 32.5%; Pred. No. 4.5e-103;

Matches 610; Conservative 179; Mismatches 530; Indels 559; Gaps 45;

QY 2 ADRIPIEAAE-DPAIVGASGRLPGGVIDLSGFYTL-EGSRDVGVRPWR-WDAAMF 58

DB 1478 AVRVGADSEDEPAIVGIGCRFPGIGSPBQLMRVLAEGANLVTG-FPADRGNDIGRLY 1536

QY 59 DDPDABGKPTVRASFLSDVACPDASFPGISPREALRMDPAHRLLEVCMEALENAIA 118

DB 1537 HPPBDNGTSTYVDKGFLLTDADBDPFGITPREALAMPQOKLMETMEANERAGID 1596

QY 119 PSAIVGTETGVTGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGRPC 178

DB 1597 PDALRGDTGVFVGMNGSQSYQULL--AGEAEKVVDGYGLGNSASVLSGRILATYTFGMEGPA 1654

QY 179 VAVDTAASSSLVAVHLACQSLRSGECTALAGCSIMLSBSTLVMLSKTALADGCKA 238

DB 1655 LTVDTASSSSLVGTHLMQALRREGCSIALAGVTAVSDPTTFVDFSTQGLASDGRCKA 1714

QY 239 FSAEADFGGEGGCAVVVLKRLSGARADGRILAVIGSAIINHGGSSGLTPVPGSSOEI 298

DB 1715 FSRADDFALSEGVALVLELISRAKANGQVLAVLGSVAVNQGASNGLAAPNGPQER 1774

QY 299 VLKRALADAGCAASSVGVGAHGTGTLGDPFIEIQALNAVYGLGRDVATPLLLGIVKTNL 358

DB 1775 VIRQALAAAGVPAADVVEAHGTGTGLDPIEAGALLIATYQDGRD--RPLRLGSVKTNL 1832

QY 359 GHPEVASGITGLKLVSLIHOQIPIAHLHAQALNPRISMGDLRLTTRARTPMDWNTPR 418

DB 1833 GHTQALAAAGVATKVLARHGLRSLHADELSPITIDMESGAVEYLREVEWPAQERR 1892

QY 419 RAGVSSFGMSGTNAHVLEEAR-----AATCTPAPERPAELLVLSARTASALDAQARL 473

DB 1893 RAGVSSFGMSGTNAHVLEEARPAEQEAARTENGPLP-----FVLSGSEAVVAQAARAL 1946

QY 474 RDHLETYPQCCGDVAFSIATTRSMENHPLAAVTSREGLRALDAAAGCOTSPGAVRSI 533

DB 1947 AEHLRTPPELGLTDAAMWTATGRARPDVRAAVLGGDRAGVCAELDLAAGRSADVAVY 2006

QY 534 ADSRQKLAFLFTGQAGOTLGMGRGLYDVMSAFREAFDLCVRLFNQELPRPREVMAPR 593

DB 2007 TSAFR-KPVLVFPGQAGQVGNARDLBSEVFAEBSMCAEALSHHTWKLLDVVRGG 2065

QY 594 ASYDALLDQTAFTQPALFTFEYALALWRSNGVEBELVAGHSIGELVAAVAVGSLEL 653

DB 2066 GPDHHRVD---VLQPVLSIMVSLAELWRAGVTTPAAVVGHSQGEIAAAHVAGALSLEA 2122

QY 654 AVFLVARGRLMOALPAGAMVSTIAPADVAANAAPHAAVSIIAAVAPDOVYITGAQ 713

DB 2123 AAKVVALRSQVRLDLDQGMVSVGSRDELETVLARMQGRVAVAAVNGPSTSVAGPTA 2182

QY 714 PYVAIAAAMAARGARTKALVSHAFSPIMAPMLLEAFGRVAVSVSRPSIVLVSNLGK 773

DB 2183 ELDEFPAEAEAEKMRRLAIVRASHSEVARELDELTALELGTITAVRSVPLHSVTVE 2242

QY 774 ACTDEVSBBQVWRHAREVVRPADGVKALHAAGAGTFVEGPKSTLL-----GLV 823

DB 2243 VIDTSMDSYWRNLRRPLFEOAVRGLVEQCFDTFEVSPHPVLMAVEETAEGAAGAE 2302

QY 824 PACMPARALLAASSAGDEPATVLEAL--GGLMAVGLVSMAGLP--SGGRRLPLRT 879

DB 2303 VTCVP-----TLREQSGPEFLRLRLRAHVGVG-----ADLRPAVAGGRPAELPT 2349

QY 880 YPMQREYV--IDTKDDAARGDRA----- 903

DB 2350 YPEHQRFPRRPRPVDVSLGVKGAEHPLLAADVDPHGCAVFTGRSTDEQPLAEH 2409

QY 904 -----PG-----AGHD----- 909

DB 2410 VVGRTLVGSLVDLALAAGEDVGLPVLBELVLAQRPVLVLAGALLRMSVGAPDESGR 2469

QY 910 ----- 909

DB 2470 TIDVHAABVDLADAOQSHATGTLAOGVAAGPRDTGEQMPEDAVRIPLDHYGLAEO 2529

QY 910 -----EV-----EEG-----GAV----- 917

DB 2530 GYEVGSFOALRAAMKDSVYAEVSIADDEBYAHPVLDDVAQTLSLGLNGERGCGK 2589

QY 918 -----RGDRRSARL----- 930

DB 2590 LPAAMTVTLHAGATSVRVATPAGADAMALRVTDPAGHLVATVDSLVRSTGEKWEOP 2649

QY 931 PPSGGR----- 947

DB 2650 EPRGEGGELHALDWGLAEBSGTGRVVAADSDLVLSRGEPEPAVLVRYEPEBDPR 2709

QY 948 ----- 947

DB 2710 AAARHGVLAALVWRKLEOEELPGATVIATSGAVTSDDSVPERGAAMMGVRCQA 2769

QY 948 -----FRLEID-EPGVL-----DHLVLRV-----TERAPG 972

DB 2770 AESPDFFVLIDTAAEKGMLPAVDPNPQALARGDVFVPRLSPLAPALTLPACTQRLVPG 2829

QY 973 -----LGEVEIAVDAAGLSFNDVOLALGMVDDDLGKKNPFLLG 1012

DB 2830 DGAIDVAEPADVDQPLRAGEVVRVDRATGVNFRDVLALLGMYPOKAD-----MG 2881

```

QY 1013 GECAGTIVAGGCVGLVVGQPVIALSAGAFATHTTSATLVLPFOALSAIEAAMPVA 1072
DB 2882 TEAAGVAVAVGPDVDAFAPGDRVLGLFGQAFPIANTDRHLARVPDGSDDAAVPIA 2941
QY 1073 YITANYALDRIRLROGGERVLIHAATGVGLAAVQMGAGVBAVHATAGTPEKRAVLESL 1132
DB 2942 YTTAAHYALHDLGLAGGOSVLLHAAAGGVMAVALARRAGAEVLAATAG-PAKHGTLRAL 3000
QY 1133 GV--RYVSDRSRDRFPAVDVAMTGGEGVAVVNSLGGELIDSEFNLRSHGFVETGKD 1190
DB 3001 GUDDEHILASSRTGFRKFRKRTGGVAVVNSLGLDLSADLADGVFVEVNGKTD 3060
QY 1191 CYADNQLGRLPFLRLNLSFSLVLDLRCGMMLEPPA-----RVRLLEELLGLLAQVF 1240
DB 3061 -----LRLDAG-----DPRG-----RYAPFDGEGADDRLGELIEVVGILGAGSL 3100
QY 1241 TPEPIATLPIAVADAFRSMAQOHLKVLVTLG---DPEVOIRLPTHAGAPSTGDRDL 1297
DB 3101 DRLPYSAWELGAPALQHMSSRGHVGLVLTOPAPVDDGTVLJ-----TGCTGT 3151
QY 1298 LDRLASAARAAALAEALFRTQVSVLTPEIKVAELFTRLGMDSLMAVELRRIETA 1357
DB 3152 LGRL-----LARHLVTEHGVRL-----LVSRRGADAGSDELRAEIE- 3190
QY 1358 SLKLTSTFTLSTPNIALLAQNLDALATL-----SLERVAEN-LR 1400
DB 3191 DUGASAEIACDTABRDALSA--LDDGLRPLTGVHAGVLADELVTSIDEPAYEQVLR 3248
QY 1401 AGVQNDVSSGADQDWEI 1418
DB 3249 AKV-----DAAMNL 3257

RESULT 2
ERY1_SACER STANDARD; PRT; 3491 AA.
ID ERY1_SACER
AC Q03131;
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
GN deoxyerythronolide B synthase 1) (DEBS 1).
OS ERYA.
OC Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiales; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Scaver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis."
RT Science 252:675-679(1991).
RN (12)
RP SEQUENCE OF 3474-3491 FROM N.A.
RX MEDLINE=93231529; PubMed=8386127;
RA Donadio S., Scaver M.J.;
RT "IS1136, an insertion element in the erythromycin gene cluster of
RT Saccharopolyspora erythraea."
RT Gene 126:147-151(1993).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- CORRECTOR: NADP, CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,

```

```

CC CC
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: Contains 3 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M63676; AAA26493.2; -
CC DR EMBL; L07626; AAA26504.1; -
CC DR PIR; T43231; T43231.
CC DR HSSP; P25715; 1MEA.
CC DR InterPro; IPR001227; Ac transferase.
CC DR InterPro; IPR002198; ADH short.
CC DR InterPro; IPR000794; Ketoacyl-synt.
CC DR InterPro; IPR006163; Pp bind.
CC DR InterPro; IPR006162; Pantane attach.
CC DR Pfam; PF00698; Acyl_transf; 3.
CC DR Pfam; PF00106; adh_short; 1.
CC DR Pfam; PF00109; ketoacyl-synt; 2.
CC DR Pfam; PF02801; ketoacyl-synt-C; 2.
CC DR Pfam; PF00550; pp-binding; 3.
CC DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
CC DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
CC DR PROSITE; PS50075; ACP_DOMAIN; 3.
CC KW Phosphopantetheine; Multifunctional enzyme.
CC FT DOMAIN 1 1972
CC FT DOMAIN 1 1979 3491
CC FT DOMAIN 1 375
CC FT DOMAIN 1 414 484
CC FT DOMAIN 1 503 961
CC FT DOMAIN 1 1030 1356
CC FT DOMAIN 1 1611 1794
CC FT DOMAIN 1 1888 1958
CC FT DOMAIN 1 1979 2441
CC FT DOMAIN 1 2507 2854
CC FT DOMAIN 1 3055 3237
CC FT DOMAIN 1 3334 3404
CC FT ACT SITE 145 145
CC FT ACT SITE 447 447
CC FT BINDING 677 677
CC FT ACT SITE 1128 1128
CC FT ACT SITE 1614 1660
CC FT BINDING 1921 1921
CC FT ACT SITE 2148 2148
CC FT ACT SITE 2598 2598
CC FT NP_BIND 3058 3104
CC FT BINDING 3367 3367
CC SQ SEQUENCE 3491 AA; 365022 MW; 682BFC32C90FA8C4 CRC64;

Query Match 26.7%; Score 1925.5; DB 1; Length 3491;
Best Local Similarity 35.6%; Pred No. 1.7e-93;
Matches 547; Conservative 184; Mismatches 601; Indels 203; Gaps 39;

QY 4 RPIEAAEDPIAIVGASCRLLPGCVTLDSGFWTLGSRDTPYGRVPAER-WDAAMPDPDP 62
DB 497 RTNEAAPPEPVAVVMACRLPGCVSTPEEFELLSEGRDAVAGLPTDRGMDDSLFFHPDP 556
QY 63 DARGTPTVTRASFSDVACFDASFCGISPRELRMDPAHRLLEFCWELLENAAIAPSL 122
DB 557 TRSGTAHORGGGFLTEATAFDPAFFGMSPRELAVDPOORLWLELSWEVLERAGIPPSL 616

```

QY	123	IGTEGVYIGIGPSPSEYEAALLPOQAFASADIDHGGIGTWPISVGARISYALDARPCAVD	182
Db	617	QASPTGVAVGLIPOEYGRPL--ABEGBEVEGLMTGTTTSTAASGRATATLLEBPAISVD	674
QY	183	TAYSSSLVAVHAIACQSLRSGECSTALAGVSLMLSPSTLVLSKTRALARDGRKAFSAE	242
Db	675	TACSSSLVAVHAIACQSLRSGESSLAMAQGVITMPTPGMLVDFSRNSSLAPGRCKAFSAG	734
QY	243	ADGGRGGCCAVVYLKRLSGARADGRLLAVITRGSALIHQDQASGLTYPNSSOGEIYUKR	302
Db	735	ANGFGMAQGMQLLERLSLDARNGHPALVALRGTAVNSDASNGLSAPNRADAVRYIQ	794
QY	303	ALADAGCAASSGVYEAHGTGTTGDPFRIEIOALNAVYGLGRDVATPLLIGSVKTNLGHPE	362
Db	795	ALABSGCLPAIDAVEAHHGTGTRGDPLEARALPEAY--GDRDROPLHSGVSKNLGHTQ	852
QY	423	SSFEGSGTNAHVLEEAR-----AATCTPPAPERPAELLVLSARTASALDQOARL	473
Db	913	SSFGISGTNAHAIIEEAPQVVEGERVEAGDVAP-----WYLSASSAAGLEAOARL	964
QY	474	RDHLETVYSOCLGDVAFSLATTRSMEHRLVAATRSREGLAALDAADAAQGTSGCAVRSI	533
Db	965	AAHLKEHGOOPRDIYATSLATRALPFRALPAPYDESALALRVLDGLTNGADGAAGT-	1023
QY	534	ADSRGKLAFLFTQCGAQTLCMGRLYVWMSAFREAPDLCYRLFQELDRPLERVMAEP	593
Db	1024	-SRAGQRAVVFPPGGQWQMGMAVDLDTSPVFAALRECADLHPHLDPEVIFPLREA	1082
QY	594	A--SYDAL-LDQTAFTOPALFTEBYALALMRSGVEBELVAGHSIGELVAACVAGVS	650
Db	1083	ARREDDALSTERVDVQVPMFAVWVSLASMWRAHGVPAVITGHSGOEIAAACVAGALS	1142
QY	651	LEDVAFVLAAGRLMQALPAGGAWSTIAPREADVAANAAPHAASVSTIAVNAPOQVYAG	710
Db	1143	LDDARVALVLSRVITATPMGNKMSIAIAPGEVARIQDR--VEIAAVGPPSVVAG	1199
QY	711	AGOPVHAATAAARAGARTKALHVSHAESHPLMAPLEAFGRVASEVSYSRRPSIV-LVSN	769
Db	1200	DSDELDRLVASTCTECITAKRLANDYASHSHSVETIRALMAHJEDPHLPGRVPPFT	1259
QY	770	LSGKACTDEVSSPGYVWRHABEVVRPADGVKALHAAGACTVEVGPKSTLLGLVPACMD	829
Db	1260	VTGRWQTPDELDAGYWMYLRRTVFAADVABALAEQGYRTELEVSAHPILTAALBEIGDG	1319
QY	830	ARPALLA--SSRAGDEPATVYLEALGILMAYGVLVSNAGL-FPSSGRGRVPLPTYPWQER	886
Db	1320	SGADISAIHSLRRGGSLADPGEALSRAPAGVADVMSYHLGTGARVRPLPTYPFORER	1379
QY	887	YWIDTKADDAARGDRRAPGAGHDEVEEGAAVR-----GGDRRSARLD-----	928
Db	1380	WMLERK--PVARRS-----TEVDEVSALRYRIEMPTGAGEPARLDGTMVLAKYAGT	1429
QY	929	HPPESSGRREKVEAAGDRPFLLEIDEPEGVLDHLVLTERRAPGLGEVE-----IAYDAA	983
Db	1430	ADETSARREALSESAGAVRELVDARGBDELA-----ERLRSVGEVAGVLSILAVDEA	1484
QY	984	GLSFVDVOLALGMVDDLP-----GKPNPPLLIGGECAGRIYAVGEGVGLV--V	1031
Db	1485	--EPEEAPALASLDLTLVQAMVSAELGCPMLMTVESAVATGFERVRNAAHQALMGV	1542
QY	1032	GOVPALASGAFAHTVTTSALVLPRLPALSIE-----AAMVAVVLTIMYALDR	1082
Db	1543	GRVITLENPAWGLVDVPAGSVAEALHLLAVVSGAGEQDLALADGVYGRRWRAAA	1602
QY	1083	IA---RLQPERVILTHAATGVGLVAANOVAOHVGAEVNA-----TAGTPEKRAYLE	1130
Db	1603	PATDDEMKPTGVVLTGGTGGVGGQILAMWLARGRP-HULLVSRSGPADAGELVABLE	1661
QY	1131	SLGVRY-----VSDSRSDRFADVAVRAMTGEQGVUVL-----NSLSGEL	1168

```

Db      1662 ALGRTTVAACDVTDRRS-----VRELLGGIGDDVPLSAVFMAATLDDGTDTLTGER   1715
        :|::|
Oy      1170 IDSFNILRSHGRFVELGKDCY-ADNQLGLRPRLRLSTSL-----VDLR   1214
        :|::|
Db      1716 IERA-----SRKVLVGARNRHETRELDLTAFLVFSSPASAAGABGLGYAPGNAYLD   1768
        :|::|
Oy      1215 GMMLEER-----PARVR-----LLEEELGLIIA-----AGFTPPP           1244
        :|::|
Db      1769 GLAQQRSDGLPATVAVMGTWASGMME---GAVALDFRRHGVTEMPEPRACALQNALD   1825
        :|::|
Oy      1245 ---IATPLPIARVADAFPSMAQAQHKLVTLLGPVEVOIRIPTHAGAPSTGDRLDLRL   1301
        :|::|
Db      1826 RAECPCIVIDVRMORFLLATYAQRPTLPDEIDAR---RAAPQAPAEPRVG-----AL   1876
        :|::|
Oy      1302 ASAAPAAPAAALEAFLETQVSQVL-RTEPEIKVGAHALFTLGMDSLMAVLRNRIEASLK   1360
        :|::|
Db      1877 ASLEPAEREELFEVLVSHAHAVALGHASASERVDPADQFAEELGYDSLALTELNRNLGAATG   1936
        :|::|
Oy      1361 LKSTFTLSTSPNIALLAQNULDALATALSLERA   1395
        :|::|
Db      1937 VRLPTTVFDHPDVRTLLAHLMALLEGATGAEQAA   1971
        :|::|

RESULT 3
OL56_STRAT          STANDARD;          PRT;    3519 AA.
AC      Q07017;
Dt      01-NOV-1997 (Rel. 35, Created)
Dt      01-NOV-1997 (Rel. 35, Last sequence update)
Dt      28-FEB-2003 (Rel. 41, Last annotation update)
De      Oleandomycin polyketide synthase, modules 5 and 6.
Gs      ORFB.
Cc      Streptomycetes antibioticus.
Cc      Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
Cc      Streptomycinae; Streptomycetaceae; Streptomycetes.
Cx      NCBI_TaxId=1890;
        [1]
Rp      SEQUENCE FROM N.A.
Ra      MEDLINE=94150470; PubMed=8107683;
Rf      Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
Rt      "Characterisation of a Streptomycetes antibiotic gene encoding a type
Rt      I polyketide synthase which has an unusual coding sequence.";
Rl      Mol. Gen. Genet. 242:358-362(1994)
RL      -|- FUDCATION MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
        -|- FACTONE RING.
Cc      LACTONE RING.
Cc      -|- COPACITOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
Cc      -|- SIMILIARITY: CONTAINS 2 acyl carrier domains.
Cc      -----
Cc      This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc      the European Bioinformatics institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.ebi.ac.uk/announce/
Cc      or send an email to license@ebi.ac.uk).
Cc      -----
Cc      EMBL; L09654; AAA19695.1; -.
DR      PIR; S43048; S43048.
DR      HSSP; P25715; IMLA.
DR      InterPro; IPRO001227; Ac-transferase.
DR      InterPro; IPRO00794; Ketoacyl-synt.
DR      InterPro; IPRO06163; Pp_bind.
DR      InterPro; IPRO06162; Pantne_attach.
DR      InterPro; IPRO01031; Thioesterase.
DR      Pfam; PF00698; Acyl_transf_2.
DR      Pfam; PF00109; ketoacyl-synt_2.
DR      Pfam; PF02801; ketoacyl-synt_C_2.
DR      Pfam; PF00550; pp-binding_2.
DR      Pfam; PF00975; Thioesterase; 1.
DR      PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR      PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR      PROSITE; PS00751; ACP_DOMAIN; 2.

```

KM Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;  
 KW Phosphopantetheine; Multifunctional enzyme; Repeat.  
 FT DOMAIN 1 3519 MODULE 5.  
 FT DOMAIN 2 501. BETA-KETOACYL SYNTHASE 1.  
 FT DOMAIN 3 569 890 ACYLTRANSFERASE (AT) 1.  
 FT DOMAIN 4 1200 1382 BETA-KETOACYL REDUCTASE 1.  
 FT DOMAIN 5 1487 1561 ACYL CARRIER (ACP) 1.  
 FT DOMAIN 6 1686 2156 BETA-KETOACYL SYNTHASE 2.  
 FT DOMAIN 7 2220 2541 ACYLTRANSFERASE (AT) 2.  
 FT DOMAIN 8 2856 3038 BETA-KETOACYL REDUCTASE 2.  
 FT DOMAIN 9 3141 3215 ACYL CARRIER (ACP) 2.  
 FT DOMAIN 10 3270 3519 THIOESTERASE.  
 FT ACT SITE 210 210 BETA-KETOACYL SYNTHASE.  
 FT NP BIND 1203 1249 ACYL-ESTER INTERMEDIATE.  
 FT BINDING 1524 1524 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT ACT SITE 1859 1859 BETA-KETOACYL SYNTHASE.  
 FT ACT SITE 2311 2311 BETA-KETOACYL SYNTHASE.  
 FT NP BIND 2859 2905 ACYL-ESTER INTERMEDIATE.  
 FT BINDING 3178 3178 NADP.  
 SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAE61F86 CRC64; PHOSPHOPANTHETHEINE (BY SIMILARITY).

Query Match 26.2%; Score 1889; DB 1; Length 3519;  
 Beel Local Similarity 34.0%; Pred. No. 1.4e-91;  
 Matches 541; Conservative 191; Mismatches 595; Indels 264; Gaps 44;

QY 9 AAEDIIATVAGSACRPGVVIDLSGFMTLESGRDTYGRVPAER-NDAAAMPDPDAGK 67  
 DB 1664 ASREIATITANSRCRGGIDSPEDIMRFLAEGRDVAGIPEDRGMDLDLHYDPDENGT 1743  
 QY 68 TPVTRASFSDVACDFASFFGISPREALMDPAHRLLEVCWEALENNAIAPSAVGTET 127  
 DB 1744 TVREGARFYDAAGDAGFGISPREALMDPQGLLLETSMELPERADIDITYTRGTAT 1803  
 QY 128 GVFTIGGSEY---EALPQATASAEIDAHGGLGMPYSGACRISYALGLRCPVAVDTA 184  
 DB 1804 GIFTAGHGGYDPDKRAPESVAGYL-----TGTAVALSGRISYTFGLGEPATVDTA 1858  
 QY 185 YSSSLVAVHLACQSLRSRGCSTALAGVSLMSPSTLWLSKTRALARGRCARFASAD 244  
 DB 1859 CSSLVLAHLVAOLARRGECSLAAGVAVMSTPAFVFRSQQMARHGRKAPAAAD 1918  
 QY 245 GFGREGCAVVVLEKLSGARADGRIILAVIRGSAINHGSAGSLTPVNGSQOEVILKRAL 304  
 DB 1919 GNGWGEVSLILLETSLDARLGHVILAVVRGSAVNGGASGLAANGPQGRVIRAL 1978  
 QY 305 ADAGCAASSVGVVEHAGTGTTLGDPITETQALNAVYGLGDVATPLLIGSVKTNLGHPEYA 364  
 DB 1979 ADAGLAPADVVEHAGTGTTLGDPITETQALNAVYGLGDVATPLLIGSVKTNLGHPEYA 2038  
 QY 365 SGITGLKLVLSLQGOIPAHLLHAQALPRISWGLRLTVTRAPRW-PDMNTTPRAVVS 423  
 DB 2039 AGVAVGMKVLNLGGVVPKTLHVDPSPHVMSAGAVELLTEERPMEERLRRAGIS 2098  
 QY 424 SEFGMGTAHVHLEAPAPATCTPPAPERPAELL-----VLSARTASALDAQAARL 474  
 DB 2099 AFGVGTNAHVIVEAPAPATCTPPAPERPAELL-----VLSARTASALDAQAARL 2154  
 QY 475 DHLETVPEQCGLDVAFSLATTSAMEHRLAVATSRBGLRAALDAAGQOTSFGAVRSIA 534  
 DB 2155 AHVSTGTAGV-DVGMSTVATRSVFEHRAVWVGTDLDSWAGSLAFAAGVAVPGVAVGA 2213  
 QY 535 DSSRGKLAFLFTGGAQTLLGMRGLYDVMSAFREAFDLCVRLFNELDRPLREVMMAEPA 594  
 DB 2214 PAEGRRVVFVFGGQSSQWMAAGLLDACPVFAEVAEBAVAALVDTVMKSLVEVQGR-- 2271  
 QY 595 SVDAALLDQTAFTQPALFTFEYALAAWRSMGVBEELVAGSHIGELVACVAGVSTLEDA 654  
 DB 2272 --DAVTLGRVDVQVQALMAVWVSLARTWRYGVVEBAVAGSHQGIATAACVAGSLDAG 2329  
 QY 655 VFLVAKRGLMALPAGGAMVSIAPBADVAANAAPHAAVSIAAVNAPDOVVIAGAGOP 714

DB 2330 ARVVLSRAIARIAGGGMWVSILPAGRVMTMLDTYGGVRSVAAVNGPSTTVSGDVQA 2389  
 QY 715 VHAIAAANAAGARTKALVHGAHFAHSPMLAEAFGRVASVYSRPSIYLVNLSGKA 774  
 DB 2390 LDELLAGRGRCVARRRVPVYVYSHSQMDQRLBELALADITQDSSVFFSVTADW 2449  
 QY 775 CTDEVSSGCVRRHAREVVRPADGKALHAAAGTFVEVGRKSTLLGLVPACMP--DARP 832  
 DB 2450 LDTTALDAGVFTMLRETVRFOEAVEGLVAGQMGAFVCSFHPVLPVPIEQTLDALDQNA 2509  
 QY 833 ALLASSRGRBEP--TVLEALGIMAVGLVSN--GLPFGSGRRVPLPTYWQER 886  
 DB 2510 AVLSLR--REBGLDRLTLTSLAEFAVGVGVVDVTHAFEGVTP--RTVDLPITYFQHR 2564  
 QY 887 YWIDTKADDAARG--DRAPGAGHDEVEEGAVR-----GGRRSARLDHPPEGRRKV 940  
 DB 2565 FWLDSPASSANGVDGEADAMIMDAVEREDSVAAEELGIDAEALHTVLPALSSRRRRV 2624  
 QY 941 E-----AAGRPFLRIDE-----PGVLDHLVLRTERAPGLGE 975  
 DB 2625 EHRRLQDRYRVEMKFPALDEVLGGMFLFVPRGLADDGVAVRVAATARG--GE 2680  
 QY 976 VEI-----AVDAAGLSPVDQALGMP--DLPKRNPPLLIGGECA 1016  
 DB 2681 VSVVELDPTDRBRARAYEAAGRGVS-----GVVSFLSMDRRHSEH--PVVPAGLAA 2731  
 QY 1017 GRIVA-----VGEVGN--GLVVGQPVYA--LSAGAFAHTVTS--ALVPRPO- 1059  
 DB 2732 SLVLAQALVDGRVGEGRMLVTRDAVAVGPSDAGVIDVQAVMGFGVGLHEPBL 2791  
 QY 1060 -----ALSAIEAA--MPVAYLTAYALDRIAL- 1086  
 DB 2792 MGLILDLPVEAPBPGSTCDHTYADLLATVVASAGBEDQVAAARGSV--VARLVRAVVDG 2849  
 QY 1087 -----QPEERVLTHAATGCVLAAYQVAQVGAEVH-----ATGTEPKRAYLESL 1132  
 DB 2850 GGGGWRPRGTVLVYGLGLGAHARTARMLVCGGAD--HVLVSRGGSSAPGADLVRELEGL 2908  
 QY 1133 GVRVYSDRSRPAVD--VRAMTG--GEGVDVVLNLS- 1166  
 DB 2909 GGARVSRAACD--VADRVALPALLSDLGEPTTAVFHAAGVQSTPLAETSVOEADVMAA 2966  
 QY 1167 -----GELID-----KSFNLRSHGRFVELGRDCA- 1193  
 DB 2967 KVAGAVNLGELVDCGLEAFVLPSSNMGWMSGGQAVYAAANAFDLALAVRRRGGLPAT 3026  
 QY 1194 -----DNQLGRPELRNLSFSLVDLRQMLERPARVALLBELIGLIAAGVFTPP 1244  
 DB 3027 SVANGMAGBGMASVGGARELS-----RRGVRAMDPERAVAVVADAVGSEAFV--A 3077  
 QY 1245 IATLPIARVADAFPSMAQOHLKVLVLTADPEVOIRI-----PTHAGAPSTGDRGL 1298  
 DB 3078 VADVWERFVTGFSANRPRLISDL-----PEVRTARNOEOQLHAPV--PEDRSACL 3130  
 QY 1299 DRLSAAPARAALAEFLRTQVSQVLRTPERIK--VGAELFTRLGMSLMAVELRNRIEA 1357  
 DB 3131 RRLMWSPAGREAEVLVLRTEAAVAVLGHSAQVPAERAKELGFSLSLTAVQLRNRLAA 3190  
 QY 1358 SLKLKLTSTFLSTSPNTALLAQNLLDALATA 1388  
 DB 3191 ATGTRLPASAVFDHPHAALRLMILLAGWRHA 3221

RESULT 4  
 ERY3\_SACER  
 ID ERY3\_SACER STANDARD; PRT; 3172 AA.  
 AC Q03133; Q54097; Q99270;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-deoxyerythronolide B synthase III) (DEBS 3).  
 GN ERYA.

```
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=91043075; PubMed=2234082;
RA Cortes J., Haydock S.F., Roberts G.A., Beville D.J., Leadlay P.F.;
RT "An unusually large multifunctional polypeptide in the erythromycin-
RL producing polyketide synthase of Saccharopolyspora erythraea.";
RL Nature 348:176-178(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Steaver W.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RL biosynthesis.";
RL Science 252:675-679(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Beville D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RL Cloning of the structural gene, sequence analysis and inferred domain
RL structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON. AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: CONTAINS 2 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X56107; CA39583.1; -
DR EMBL, M63677; AAA26495.1; -
DR EMBL, X62569; CAA4449.1; -
DR PIR, S13595; S13595.
DR PIR, S22012; S22012.
DR HSSP, P00101; ICCH.
DR InterPro, IPR001227; Ac_transferase.
DR InterPro, IPR002198; ADH_short.
DR InterPro, IPR000794; ketoacyl-synt.
DR InterPro, IPR006163; Pp_bind.
DR InterPro, IPR006162; Pantan. attach.
DR InterPro, IPR001031; Thioesterase.
DR Pfam, PF00698; Acyl_transf. 2.
DR Pfam, PF00106; adh_short. 1.
DR Pfam, PF00109; ketoacyl-synt. 2.
DR Pfam, PF02801; ketoacyl-synt_C. 2.
DR Pfam, PF00550; pp-binding. 2.
DR Pfam, PF00975; Thioesterase. 1.
DR PROSITE, PS00012; PHOSPHOPANTETHEINE. 2.
DR PROSITE, PS00606; B_KETOACYL_SYNTHASE. 2.
DR PROSITE, PS50075; ACP DOMAIN. 2.
DR TRANSFERASE; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW Phosphopantetheine; Multifunctional enzyme.
FT DOMAIN 1 1484
FT DOMAIN 1485 3172
FT DOMAIN 317 484
FT DOMAIN 554 878
FT DOMAIN 1116 1298
FT DOMAIN 1394 1464
FT DOMAIN 1488 1954
FT DOMAIN 2021 2335
FT DOMAIN 2555 2735
FT DOMAIN 2821 2891
FT DOMAIN 2926 3172
FT ACT SITE 199 199
FT ACT SITE 643 643
FT NP BIND 1118 1164
FT BINDING 1427 1427
FT ACT SITE 1661 1661
FT ACT SITE 2112 2112
FT NP BIND 2557 2605
FT BINDING 2854 2854
FT CONFLICT 231 231
FT CONFLICT 240 240
FT CONFLICT 289 289
FT CONFLICT 493 493
FT CONFLICT 493 517
FT CONFLICT 510 510
FT CONFLICT 513 513
FT CONFLICT 525 525
FT CONFLICT 536 536
FT CONFLICT 547 551
FT CONFLICT 553 553
FT CONFLICT 673 673
FT CONFLICT 716 716
FT CONFLICT 734 736
FT CONFLICT 896 896
FT CONFLICT 896 896
FT CONFLICT 988 994
FT CONFLICT 1108 1116
FT CONFLICT 1124 1126
FT CONFLICT 1132 1132
FT CONFLICT 1192 1192
FT CONFLICT 1194 1194
FT CONFLICT 1277 1278
FT CONFLICT 1385 1390
FT CONFLICT 1485 1485
FT CONFLICT 1518 1518
FT CONFLICT 1601 1601
FT CONFLICT 1724 1725
FT CONFLICT 1732 1732
FT CONFLICT 1732 1732
FT CONFLICT 1739 1743
FT CONFLICT 1762 1762
FT CONFLICT 2252 2252
FT CONFLICT 2275 2277
FT CONFLICT 2408 2408
FT CONFLICT 2420 2421
FT CONFLICT 2443 2444
FT CONFLICT 2596 2596
FT CONFLICT 2609 2609
FT CONFLICT 2715 2722
FT CONFLICT 2754 2754
FT CONFLICT 3172 AA; 331474 MM; DBBD5094E77DD5F CRC64;
SQ SEQUENCE
Query Match 26.0%; Score 1875; DB 1; Length 3172;
Best Local Similarity 34.6%; Pred. No. 6; 9e-91;
Matches 549; Conservative 184; Mismatches 602; Indels 254; Gaps 48;
```

QY 4 RPIERAEDPIAIVGASCRIPGVIDLSGFMTLLSSRDTVGRVPAERMDAAAMPDDPD 63  
 DB 31 REVEHAGEPIAIVGACRFPDGDVSPESEFVSGGDAIAEAPDR-----CM-EPDD 85  
 QY 64 AROKTVTAASPLSDVACPDASFFGISPRPALMDPAHLLLEVCHEALENAIAASALV 123  
 DB 86 A-----RUGMIAAAGDPDAGFFGISPREALAMDPOQRIMLEISWEALERACHDVSUR 139  
 QY 124 GTEITVFIGIGPESEYEAALPOATASAEIDAHGGLGTMPSVAGRISYVALGRPCVAVNT 183  
 DB 140 GSAITVFTGVGTVDYPRPDE--APDEVLCYVGTGTASSVAGSRVAVCYGLEBPANTVDT 197  
 QY 184 AVSSSLVAHVLACQSLRSGECSTALAGVSLMLSPSTLWLSKTRALARDGRCKASAA 243  
 DB 198 ACSSGGLTALHLMESLRDECGALAGVTVMSPGAFTFRSQGLAADGRCKPSKAA 257  
 QY 244 DGFREGCAVYVVKLSGABADGDRILAVIRGSAINHOGASGLTPVNGSSSEIYLKA 303  
 DB 258 DGFGLAEGAGVVLQRLSARREGRPVLAVLAGSANOQASNGLTAPSGAQVIRRA 317  
 QY 304 LADGCAASSVGYEAHGTGTTGDDPIEIOALNAVYGLGRDVAITPLIGSVKTNLGHPEX 363  
 DB 318 LBNAGVRAGDVYVEAHGTGTRIGDPIEVHALLSTYGAERDPDDPLMIGSVSNIGHTQA 377  
 QY 364 ASGITGLLVLSLOHGOIPAILHAQALNPRISWGDRLTVTRARTPMDPMTPRAGVS 423  
 DB 378 AAGVAGKAVLALRHEMPRTLHPDESPQIEMDCAVSVGQASWPAERPRAGVS 437  
 QY 424 SFGMSTNAHVLEEARPAATCTPRAPERPAELLVLSARTASALDQAARLDHLETPQ 483  
 DB 438 SFGISTNAHVLEEARPADEPPADSGPVLTVLSGRDQANRAQAGRLADLAEPEPN 497  
 QY 484 CIGDVAFSLATFSAMEHRLAVATSR-----GLRALDAAAGQTSPGAVRISADSRK 540  
 DB 498 SLRDIGFTLATRASAMEHRAVVGDDDEALAGRAVADRRIADRTITQO-----GPNSPR 553  
 QY 541 LAFLFTGGAGQTLGMGRGLYDVMSAFREAFDLCVRLFNOLDRPREVN--WAEPAVDAA 599  
 DB 554 VAMVFPQGAQWQGMARDLRESQVADSIRODCERALAPHVDSITDLISGARP----- 607  
 QY 600 LDDQTFPTPALFTFEYALALMRSNGVPELVAGHSIGELVACVAGPFLSDAVFLVA 659  
 DB 608 -LDRVDVVPALFAVWVSLAALMRSHGVBEPAVAVGSHOEIAAHVAGALTLEDAKLVAA 666  
 QY 660 ARGRLMQLPAGGAWVSEAPADVAAVAAPHAASVIAVANP--DQVVIAGAGCVHAI 718  
 DB 667 VASRVIRRLGGCGGMAFGLGTGQAAERIGRFAGALSTIASVNGPVSVVVAGSGPLDEL 726  
 QY 719 AAMAAARGARTKALHVSFAHSPMLAMPLEAFGRVAESVYRRPSIVLSNLSGKACTDE 778  
 DB 727 IAECEAEAKKARIIPVDVASHSPQVESLREBELTELAGISPVASADVALVSTTGGQPIDTA 786  
 QY 779 VSSPGVWZHAREVVPADGVKALHAAGCTFVEVGPCKSTLGLVACMPDARPA----- 833  
 DB 787 TMDTAVMYANLREOVAFODATROLACAGFAFEVBPVPLTVGIEATLIDSLAPADAGAC 846  
 QY 834 LLAASSRAGDEPTVLALGLMAVGLVSWAGLFPSSGRGRVPLPTPYMORERWIDTKA 893  
 DB 847 VGTGLRDRGGLADFTALGEBAVAGVEVDWSPAF--ADAPVELVYPPO--RYWLPITP 903  
 QY 894 DDAAR---GDRAPGAGHDEVEGEAVRGDRRSARLDHP--PPE--SGRREVEAAGDR 946  
 DB 904 GGRARDEDDMRQVVMREAMEWSASLAG---RVLLVTGCVSELSDAIRSLESGANT 960  
 QY 947 PERLEID-----EFGVLHL---VLRVTERA--PGIGVEI--AVDAAG----- 984  
 DB 961 VLTCDVESSTIGTALAADTALSTVGAAPPHGAVPSPDLALVALGAAAGVADL 1020  
 QY 985 --LSFNDVOLALGMVDDLPKGNPPLLLGEGECAGRIVAG-----GVNGVLVAGPVIA 1037  
 DB 1021 WULTRAAVAVADELVLD--PAQ---AMVGG--LGVVGIEDQGRGGLVDLV----- 1065  
 QY 1038 LSAGAPATHVTTSAALV--PRPOLSAIEAAMPVAVLTAWVALDRIARLP--GERVLH 1095

DB 1066 ---DADAASIRSLAAVLAADPRGEEQVAIRADGIKVARIIVAPARARTHLEPLAGVLT 1122  
 QY 1096 AATGVGLAAVQWAGHVAEYVA-----TACTPEGRATLESIGRVY-----VSS 1140  
 DB 1123 GGTGTGIGHALWLRSGAE--HLVILGRRGADAPAGSELRBELTALGCGVTIAACDVADR 1181  
 QY 1141 RSDRFVADVRATGEGVDVVLN---SLSGELIKFSFNILRSHGRFVELGKRDVYADNQ 1196  
 DB 1182 ARLVLAALAAAEGRITVSAVHAAGVSTSTLDD-----LTAETFEI-----AD-- 1227  
 QY 1197 LGLRPLRLNLSFLVDLRGMLEPRARVALLLELLGLIA---AGVTPPIATLPIAR 1252  
 DB 1228 -----YKVGVTV-----NLDELCPDIDAFVLFSSNAGVMSPGIASVAAAN 1268  
 QY 1253 V-ADAFRMAQA-----OHGKVLVTLGDP-----EQOIRI 1282  
 DB 1269 AFLDGPAPARAREGAPVTSIANGWAGQNMAGDEGGEYLRSGLRAMDRAVEELHTL 1328  
 QY 1283 PTHAGAGSTGDRD-----LIDLASA-----APARAAA----- 1312  
 DB 1329 -DHGQTSVSVMDRRFVLELFTARHRPLPBEIGARAREAROSEBGRPALAQLALCD 1387  
 QY 1313 -----LEAFLETOVSQVL--RTPEIKVGAELFTRLGMSLMAVELRNRIEASLKLST 1365  
 DB 1388 GREHEHLALIRAEVAAVVGHDGDAIDRDRAFRDLGFSMTAVDLRRLAIVTGVEAA 1447  
 QY 1366 TPLSTSPNIALAQLDLALATLRLERVAE-----NLRAQVN 1405  
 DB 1448 TVFDPHPTTLRDLHYLERLVGAABEAQAPALVREVPKADDPFIAVGMACRFPQGVN 1507  
 QY 1406 -----DFVSSG-----ADDMETIAL 1421  
 DB 1508 PGEIWEFTVGGSDATVEMPTRGMDLAL 1536

RESULT 5  
 MCAS MYCBO  
 ID MCAS MYCBO STANDARD; PRT; 2110 AA.  
 AC 002251;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Mycobacteric acid synthase.  
 GN MAS.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_Taxid=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=92406887; PubMed=1527058;  
 RA Matur M., Koliatukudy P.E.;  
 RT "Molecular cloning and sequencing of the gene for mycobacteric acid  
 synthase, a novel fatty acid elongating multifunctional enzyme, from  
 RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guérin.",  
 RL J. Biol. Chem. 267:19388-19395(1992)  
 CC - FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH  
 CC METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO  
 CC FORM MYCOCEROSYL LIPIDS.  
 CC - COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE.  
 CC - SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL  
 CC ARRANGEMENT.  
 CC - SUBCELLULAR LOCATION: Membrane-associated.  
 CC - SIMILARITY: Contains 1 acyl carrier domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```

CC -----
DR EMBL: M95808: AAA25369.1: -
DR PIR: B44110: B44110.
DR HSSP: P73283: 1E5M.
DR InterPro: IPR001227: Ac transferase.
DR InterPro: IPR002085: Adh zn family.
DR InterPro: IPR000794: ketoacyl-synt.
DR InterPro: IPR006163: PP bind.
DR InterPro: IPR006162: Ppanne attach.
DR Pfam: PF00698: Acyl_transferase.
DR Pfam: PF00107: ADH_zinc_N_1.
DR Pfam: PF00109: ketoacyl-synt; 1.
DR Pfam: PF02801: ketoacyl-synt C; 1.
DR Pfam: PF00550: PP-binding; 1.
DR PROSITE: PS00012: PHOSPHOPANTHETHEINE; 1.
DR PROSITE: PS00066: B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS50075: ACP_DOMAIN; 1.
KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
Transferrase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT DOMAIN 1 430 BETA-KETOACYL SYNTHASE.
FT DOMAIN 533 852 ACYL TRANSFERASE.
FT DOMAIN ? ? ENOYL REDUCTASE.
FT DOMAIN ? ? BETA-KETOACYL REDUCTASE.
FT DOMAIN 2026 2096 ACYL CARRIER (ACP).
FT ACT_SITE 177 177 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 623 623 ACYL TRANSFERASES (BY SIMILARITY).
FT NP_BIND 1561 1578 NADP (BR).
FT NP_BIND 1765 1780 NADP (KR).
FT BINDING 2059 2059 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 2110 AA; 225577 MW; C658215D7155300 CRC64;

```

```

Query Match 25.7%; Score 1852.5; DB 1; Length 2110;
Best Local Similarity 29.5%; Pred. No. 6; Se-90;
Matches 52; Conservative 176; Mismatches 519; Indels 555; Gaps 30;

```

```

QY 13 PLIAVGSARLPFGVIDLSGFMTLLSGSRDTVGRVPAERWDAAMFDDPDAPGKTPVTR 72
DB 7 PVAIVGICGRLPGINSPPDKLMESLRGDDLVTEIPDRWDADYDPEGVPCRSVSRW 66
QY 73 ASFLSDVACPDASPFGISPREALRMDPAHLLLEVCHEALENMAIASALVGTETGFTG 132
DB 67 GGFLLDVAGFPAEFPFISEREATSIDPOQLLETSEALIEHAGLDASASSTAVFTG 126
QY 133 IGSEVEAALPQATASAEIDAHGGL-----GTMPSVAGARISYALGKGPVAVDTAY 185
DB 127 LTHEDYIVLTTHA-----GGLASPYVVTGLNNSVASGRILAHITGLHGPMTPTTAC 177
QY 186 SSSILVAVHLACQSLRSGECSTALAGVSLMLSPSTLVMLSKTRALARDGCKAFSAEDG 245
DB 178 SSGIMAVHLACRSIHDERDLALAGCAVLLLEPHACVAASAQGLMSTGRCHSPDADAG 237
QY 246 FGRGEGCAVVLKRLSARADGRIILAVIRGSAIHHGASGLTVPNGSOEIVLKALA 305
DB 238 FVSEGGCAVVLKRLPALRDGNRIFAVVRGTATNQGRTETLTMPSEDAQVAAYRALA 297
QY 306 DACCAASSVGVENHGTGTLGDPRIEIOALNAVYGLRDVATPLLISSVKNLGHPEYAS 365
DB 298 AAGVQPTVGVEHGHGTPIGDPIEYRSILARYGAG-----TPALGSAKSNMCHSTASA 353
QY 366 GITGLKVLVLSLOHGQIPAHILHAQALNPRISWGDRLTTPRAATPMDWN--TPRAAGV 423
DB 354 GTVGLIKAILSLRGGVVPRLHFNRLLPDELSDVETGLFVPOAVTPMNGMDHTPKVAVS 413
QY 424 SFGSGSNNAHVLEEAALATCTPPAPRRPAE-----LLVLSAKTASALDAQAARLHLET 479
DB 414 SFGSGSNNAHVLEEAAPAEASAPESPQDAEVPRLFMSTSSDALARQTAQLATVVEE 473
QY 480 YPSQCL--GVAFSLATTRSAMERHLVAATSR-----EGLRPAALDAAGQTSFGAARSI 533
DB 474 H-QDCVAASDLATYTLARGRRRRPVRTAVVANAIPELVEGLREVADGALIDAAG----- 527
QY 534 ADSRRGLAFLPTGGAGTILGMSGGLYDVNSAFREAFDLCVRLFNQELDRPLREVMMAEP 593

```

```

DB 528 -HGDRGPV-WVPSGQGSQWAMGTOLL-----ASEPVFAATI-----AKLEPVIAES 573
QY 594 A-SVDAALDDQTAFT-----OPALFTFEYALAA-LWRSWGEPELVAGHSIGELVAACYA 646
DB 574 GRSVTEALITRAQCTVTIDKVPVFAVVOVALAINTMEQTVGRPGAIVGSMESAAVYA 633
QY 647 GVFSLEDAVELVAARGRLMOALPAGAVNSIEAREADV-AAAVAPHAASVSIAAVAPDQ 705
DB 634 GALSLEDAARVICRSKSLMTRIALAGAMGSVELPAKVNSSELMARQIDDVVSVVASPOS 693
QY 706 VVIAGAGQVYHAIIAAAMARGATKLVHSHAFHSLAMPMEFAFRVAVESVSRPSIV 765
DB 694 TVVGSTSDYVRDLIARBEQRDVAREAVADVASHSPQVDPDIDDLAALADLAIPMTKPV 753
QY 766 LVSNLSGKACTDEVSFGYVVRHAREVREPADCVKALHAAGATFEVGPCKSTLLGLVA 825
DB 754 YISATLFPDRRQPCCGAYVWDLNRTVTPFAAAVQAAEMDGIRVFADVSPHLLTAAVQ 813
QY 826 C-MPDARPALLAASRAGDEPATVLEALGLMAVGLVSWAGLFPSSGRRVPLPTYMQ 883
DB 814 TGRSLMSVAALAGMRREQPLPHGLGLLTLEHRAAGALDYSALYPA-GRLVDAPLPAWT 872
QY 884 RERYWID----- 890
DB 873 HAFLLFIDDDGQEQRAQACTITVHPILGSHVRLTEPERHWWQGVGTSVLSWLSQYH 932
QY 891 ----- 890
DB 933 NVAALFGAAVCEMAALAAAEVGEAAEVDITFEQMLLDEQTPIDAVASIDAPGVNFT 992
QY 891 -----TKADDAARGDRAPG----- 905
DB 993 VETNRDGETTRHATATLRAAEDDCPPGYDITALLQHPHAVNGTAMRESFAERGVTLGA 1052
QY 906 -----AGH-- 908
DB 1053 AFGGLTAAHTPRRPRRCWPRSRCPRRSGSSRAPTESTRCWTLVSSRARASRYGHW 1112
QY 909 -----DEVEEGG--AVRG-- 919
DB 1113 PAVAVGCAOPACLRAYPQCPRLHAVDGLQRDPWEADLVLDHGVTLAVALKRLRG 1172
QY 920 -----GDR--SARLDHP----- 930
DB 1173 TGTSEDERDRLVSERLTLTGMOQRALPEVGDEAGSWLLIDPSNAVDPDMLASTLTDA 1232
QY 931 ----- 930
DB 1233 LKSHGQGTGECASLWSVODTPPNDAGLEKLSQLRGDDGVIVYGVPRVGDDEHSLIA 1292
QY 931 -----PPESGRREKVEAAGDR----- 946
DB 1293 GREQVHLVRIITRELAEPFEGELPRLFTVTRQAVIVPHSGERANLEQAGRLRLAVIS 1352
QY 947 ----- 946
DB 1353 EHPMLRTLLIDVDEHTDVERVAQQLLSGSEDEETARNGDWYVARLTPLPGLHEERTAV 1412
QY 947 -----PFRLEIDEPVLDHLVLRTERARAPGLGEVIEIVADAAGLSFNDVOLALGNVP-- 998
DB 1413 LDDHDGMRQVARRPDDLQTLFVNASDRVPPGGQIEVAVSMSSINIFADVLIAFGRPPII 1472
QY 999 DDLPGKPNPPLLGGECAGRIIVAVGCVNGLVVGGCVIALS-AGAFATHTVTSAAVLPR 1057
DB 1473 DD--REPO-----LQMPFVGVTYAVGCVTGHQVGDVGVGFSBEGCKRTPLTCANLAVTL 1526
QY 1058 POALSAIEAAMPVAVLTATYALDRIARLOPGERVLIIHAATGCVGLAAVQMAQVGAEVH 1117
DB 1527 PRLTLTDEQATITATTAHTATWYGLNDLAQIKAGDKVLIHSATGCVGOAALISARAKAEIR 1586
QY 1118 ATAGTEKRAYLESIGVRYSDRSRPFVADVAMTGGEGCVNVVLSLSEGLIDKSPNLL 1177
DB 1587 ATAGNPAKRAMLRDMGVHEHYDSRVSVEFAEQIRRDITDYGVDIVLNSLTGAAGRAGLELL 1646

```

QY 1178 RSHGRFVELGKRDVADNOLGAPFRNLSFSLVDRGMWLERPARVALLLELLGLIA 1237  
 DB 1647 PFCGRFVEIGKADVGNTRGLGFPFRGLTFYLLDLALMSVTPDRVRELLATVFLTLTD 1706  
 QY 1238 GVFTTPEPTIPLRIARVADAFRMAQOHGKVL 1271  
 DB 1707 GVLTAAQCTHYPLAADAIRAMSNABHGTJVL 1740

RESULT 6  
 PSB\_MYCTU STANDARD; PRT; 1538 AA.  
 ID PSB\_MYCTU Q10978; OS3234;  
 AC Q10978; OS3234;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Phenolphthalein synthesis polyketide synthase ppsb.  
 GN PPSB OR RV2932 OR MT3002 OR MTCY338.21 OR MT0011.01.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;  
 OC Corynebacteriales; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Broesch R., Parkhill J., Garnier L., Church G.E. III, Tekle E.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teka E.,  
 RA Badcock K., Baaham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Suleon J.E., Taylor K., Whitehead S., Barrell B.G.,  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHEICEROL  
 CC SYNTHESIS.  
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHEINE (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: Contains 1 acyl carrier domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL021070; CA15929.1; -  
 DR EMBL; AF007122; AAK47329.1; -  
 DR FIR; E70874; E70874.  
 DR TIGR; MT3002; -  
 DR TubercuList; RV2932; -  
 DR InterPro; IPR001227; Ac transferase.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR006163; Pp\_bind.  
 DR InterPro; IPR006162; Ppantn\_attach.  
 DR Pfam; PF00698; Acyl\_transf; 1.

DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt C; 1.  
 DR Pfam; PF00550; pp-binding; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTHEINE; FALSE\_NEG.  
 DR PROSITE; PS00075; ACP\_DOMAIN; 1.  
 DR Hypothetical protein; Multifunctional enzyme; Oxidoreductase;  
 KW transferase; NADP; Phosphopantetheine; Complete proteome.  
 FT DOMAIN 1458 1495  
 FT BINDING 1458 1458  
 FT SEQUENCE 1538 AA; 162527 MW; B5E2A2042AD00CC CRC64;  
 Query Match 22.1%; Score 1592.5; DB 1; Length 1538;  
 Best Local Similarity 31.2%; Pred. No. 2.2e-76;  
 Matches 488; Conservative 207; Mismatches 600; Indels 271; Gaps 44;  
 QY 3 DPEPIBAEDPPLAIGASGRPLPGVYIDSGFTLLLEGSMDTGRPAEMNDAAAFDPP 62  
 DB 25 DVSRIAVAEPVAVVIGCFRPGDVGPEFDFVAGNNAISTVPADMDAEAFYHPD 84  
 QY 63 DAPGKTPVTRASFSLDVAFCFDSFPGISPREALRMDPAHRLLEVCWALENAAIAPSA 122  
 DB 85 LTFGKMTTGWGFVDVAGFDFAFPGITPREAANDPQGRMLLEVAEMLEHAGLPPDL 144  
 QY 123 VGTETGVFIGPSEYEALPQATASABIDAAGLGTWPSVAGRIYVALGLRCPVAVD 182  
 DB 145 GGTIRIAVMGVFVNEYSWL--AASPQNVDAVSGTGNASITVGRISYLLGLRGPVAVD 202  
 QY 183 TAYSSIVAHLACGSLSGESTLACGVSMLSPSTLWMLSKRALDRCAFSAE 242  
 DB 203 TACSSSLVAHLACGSLRRETDLAGVSTLLEPEQIATSAAGLSPQRCFAFDA 262  
 QY 243 AGFPRGECVAVVVKRLSGARADDRILAIRGSAIINHDSGSLTVNGSSQEVLR 302  
 DB 263 ADFVRGEGAGVVLKRLTDVARDQVLAIVRGSANVODGNSNGVTAENTAAQCCVIND 322  
 QY 303 ALADAGCAASVGVYEAAGTGTLLDPIEQALNAVYIGLRDVAATPLIGSVKTNLGP 362  
 DB 323 ALRSGDVAPDSVNYVEAHGTGLDPIEFALATYGHGDA--LGAVKTNIGHLE 379  
 QY 363 YASGITGLKLVLSIQHQIIPAHLAQALNPISMDRLTTRARTPPDNTRRAGV 422  
 DB 380 AAGAGIAGFKATLAQORATIPNLHFSQNPIDAASRFPPTQNSPPTREGRRRAV 439  
 QY 423 SFGSGTAAHVLE--PAACTCPAPERPELLVLSARTSALDAQARLDHLETP 481  
 DB 440 SFGAGTAAHVITIQGSELAVSSEGEPTGVTLVGTAKTORAAATQVLADMEGEG 499  
 QY 482 SQ-CIGDVAFLATRSAMEHRLAVALAATRSREGRLAALDAAGQTSPGAIVSIADSRCK 540  
 DB 500 AEVAVADVAVHTNNHRAQATGTIVARDRAQAIAGLALAAAGQAPGV--SHQGSPP 558  
 QY 541 -LAPFTGQAGTLCMGRLYDVMSAPREAFDLCVRLFNQELDRPLREYMAEPASVDA 599  
 DB 559 GTVFYVSGSGSQWAMGRQLADDEPAFAAVALPEPVEQAGFSLRDI--ATGKELY 615  
 QY 600 LLDQTAFOPLFTFEVYLAALMRSWGVEPELVASHSIGELVAAACVAGFSLDVAFLVA 659  
 DB 616 GIEQ--IQGLIGMQLTTLTLMRSYGVQPDVLIHSGEVAVAAGLALTAEGRRVTA 672  
 QY 660 AGRGLMQLPAGAMVSIIEAPADVAAVAPHAASVSIAAVAPQVIVAGAPVHAIA 719  
 DB 673 TRARIMAPLSGGGMAALGLDAAATEALIDV--PQVTGIVNSPQGTIVAGTBEQIDELI 731  
 QY 720 AAMAAKGATKALHVAHSHFSLAPMLIEAFGRVAVESYRPPSIVLVNSLGGACTDEV 779  
 DB 732 AEVRAQNFASRVNIEVAPHPNDALQPAKMSLEADLTPRTPTGIISTVYADLHTOP 791  
 QY 780 SSGVWVHAREVAFADGVKALHAAG-----TFVEVGKSTLLGLVPCMPDA--RP 832  
 DB 792 PDAEHWATNMRNPVFO--QAIASAGGADGAYTFTEISAPLITQAIADTLEDAHP 848  
 QY 833 A-----LLASPRAGDEPATVLEALGLMANVGLVSMAGLPPSGGRVP----- 876

```

Db      849 TRSAKYLISICGLORDADTV-----TFRNTLYTADIAHPHTCHPBPBHT 895
Qy      877 LPTVPMQREKRYNIDTKADDAARGRAPGAGHDEVEEGAVRGDRSALDHP---PPE 933
Db      896 IPTTPMOTHMWTAT-----THPSTAPE 919
Qy      934 SGREKVEAAGDR-----PFR--LEIDEPGV-----LDHLVLR 965
Db      920 DPGSNKYVNVGCGSTSESRALDEDMCHOLAMPTRAVSDPPSTAMLVADNLCHELARA 979
Qy      966 TERRAPGLGEVEIAVDAAGLSPNDVOLALGMPDPLGKPNPPLLGGECAGRIAVAGEG 1025
Db      980 ADSRVDLSPPALAGSDPALLD---ALRGVDNVLVAPVPGLLDIESAYGVFHATRR 1036
Qy      1026 VNGLVVQPIVIALSAGFAFHVTTSA-----ALVLPRLQALAI 1064
Db      1037 LAAMVAASATATSPKLTFT-MTRNMQPISEGRANDGHAVLWGRLSLLEHETWGI 1095
Qy      1065 -----EAAAMPVAVLTAWVALD-----RIARLO-----PGERVLIHA--- 1096
Db      1096 IDLDDSPALALAVRHV---LTHAGTDBEDQVYVRSGRAPVRLQRTLPKRPVTLNADAS 1153
Qy      1097 -----ATGVGLAAVQAOHVGAE-VHATAGTP-----EKRAYLESIGVRYS----- 1138
Db      1154 QLVYATGNIGPHILIRQIARHGAKTIVAMARKPGALDELQCLAAATGDLIAVAADATDP 1213
Qy      1139 ---DSRSDRVADVVRATG-----GGEVDVVLNSLSELDKSN-----LNSHGFEVE 1185
Db      1214 AAMQTLFDRFGTELPLEGIIYLAFAQRPALISEMTDDVTWFRPDLALALHRSLSK 1273
Qy      1186 -----LGKR-----DCYADNOLGPRFLR-----NLSPFLVDLR 1214
Db      1274 SPVRHFLVSSVGLSGRLAHATATSAFLDSAGARRIMGRLPATVDMGLMSLADVO 1333
Qy      1215 --GGMTERPARVALLTEL-LGLIAGVFTPPPIATLPAR---VADAFRMAQAOHLG 1267
Db      1334 KDATQISAESGLQMADEVAIGALPLVMNPDAVAATVVAADWPLAAAVRTGALRIVD 1393
Qy      1268 KLVLT---LGDPEVQIPIPHAGSGPSTGRDL-DLAAAPAAARAAALAEAPLRTVQSO 1323
Db      1394 DLLPAPEDVGKGESEFR--TSLSRCPAEKRDMLFDHVGLAATWVG----- 1438
Qy      1324 VLRTPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTPTLSTSPNIALAONLID 1383
Db      1439 --MPPTPLDPSAGFQLGNDLSMSTVLQALSLSELPFAPASVVPYPTV---YSLTD 1492
Qy      1384 ALATATL 1389
Db      1493 YLATVTL 1498

RESULT 7
PPSA MYCTU STANDARD; PRT; 1876 AA.
AC 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenolphthalein synthesis polyketide synthase ppsa.
GN PPSA OR RV2931 OR MT3000 OR MTCY338.20.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1773;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eismeler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,

```

```

RA Hornsby T., Jagele K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sleson J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Bishai W.;
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPTHCEROL
CC SYNTHESIS.
CC -1- COPACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES (By
CC similarity). Contains 2 acyl carrier domains.
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z74697; CA98988.1; -.
CC EMBL: AE007122; AA847328.1; -.
CC PIR: C70749; C70749.
CC DR TIGR: MT3000; -.
CC DR Tuberculist; RV2931; -.
CC DR InterPro: IPR001227; Ac transferase.
CC DR InterPro: IPR000794; Ketoacyl-synt.
CC DR InterPro: IPR006163; Pp-bind.
CC DR InterPro: IPR006162; Pantine_attach.
CC DR Pfam: PF00698; Acyl_transf. 1.
CC DR Pfam: PF00109; ketoacyl-synt. C; 1.
CC DR Pfam: PF02801; ketoacyl-synt. C; 1.
CC DR Pfam: PF00550; Pp-binding; 2.
CC DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
CC DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
CC DR PROSITE: PS50075; ACP_DOMAIN; 2.
CC KM Multifunctional enzyme; Oxidoreductase; Transferase; NADP; Repeat;
CC KM Phosphopantetheine; Complete proteome.
CC FT DOMAIN 7 80 ACYL CARRIER (ACP) 1.
CC FT NP_BIND 1498 1503 ACYL CARRIER (ACP) 2.
CC FT BINDING 62 62 NADP (POTENTIAL).
CC FT ACT_SITE 273 273 PHOSPHOPANTETHEINE (POTENTIAL).
CC FT ACT_SITE 720 720 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC FT BINDING 1796 1796 MALONYLTRANSFERASE (BY SIMILARITY).
CC FT BINDING 624 624 PHOSPHOPANTETHEINE (POTENTIAL).
CC FT CONFLICT 877 877 D -> E (IN REF. 2).
CC FT CONFLICT 877 877 R -> H (IN REF. 2).
CC FT CONFLICT 1323 1323 G -> S (IN REF. 2).
CC SQ SEQUENCE 1876 AA; 198834 MW; D9783DBD48792110 CRC64;

Query Match 22.0%; Score 1583; DB 1; Length 1876;
Best Local Similarity 29.2%; Pred. No. 8,9e-76;
Matches 532; Conservative 205; Mismatches 622; Indels 460; Gaps 49;

Qy 2 ADRPIERAA---EDPIAVGASCRLLPGVYIDISGFWTLLEGSDTGVGRVPAERMDAAW 57
Db 88 SDAAVKRGANSLDEBIAVAVGMCRRPGGISCEALMDFLCERRSSISQVPPORWOPFEG 147
Qy 58 FDDPDPAKPTPTVTRASFSDVACPDASFPGISPPREALMDPAHRLLTEWCALNMAI 117
Db 148 GPPEVAAALARTTRMSFLPDIDAFDAEFELSPSDAKMDPOORLLLEVAAWALSHAGI 207

```

QY 118 ABSALVGTETGVFIGBSEYEALPOATAS--AEIDAHGLGTMSVGAGRISALGLRG 176  
 Db 208 PGTLRSATGVFAGACISEYGA---MASADLSQVDMGMSGGAMSIIINRLSTFLDLRG 264  
 QY 177 PCVAVDTAVSSSLVAVHLACQSLRSGECTALAGVSLMSPSTLVWLSKTRALADGRC 236  
 Db 265 PSVAVDTACSSSLVHILACOSLRQDCHIAIAGVNLISPAVFGFQVQLASPTQC 324  
 QY 227 KAFSAADGFGGEGCAVVVLRKLSGARADGRILLAVIRGSAIINHGASGLTVPNGSSQ 296  
 Db 325 RAFDITADGFGVBEAGVAVVLRKLTDAQRDGRVLAIVGSAVNOQGRSNGMLAPRPAQ 384  
 QY 297 ELVLKRALADAGCAASVGVYEAHGTGTGADPIETIOALNAYYGLGRDVPATLLIGSVT 356  
 Db 365 MAVLRAAYTNAGMOQSEVDVYEAHGTGTLGDPTEARLGTVLGRGPRPDSLLIGSVT 444  
 QY 357 NIGHEBYASGITGLKVLSLQHGQI PAHLHAQALNPRISWQDLRLTVTRATPMPDMNT 416  
 Db 445 NIGHEBAAGIAGFKTIVLAVOHQI PRNOHFETANPHI PFTDLMKVVDQTEMPATGH 504  
 QY 417 PRAGVSSFGMSGTNAHVLEEARATCTPRAPERPA--ELVLSARTASALDAQARLD 475  
 Db 505 PRAGVSSFGMSGTNAHVLEEARATCTPRAPERPA--ELVLSARTASALDAQARLD 564  
 QY 476 HLETPSQ--CLGDVAFSLATTRSAMEHRLAVATSRREGRLAALDAAQOTSPGAVRSIA 534  
 Db 565 WMEGADVALADVAHTLNHRSRQKPGFTVVARBRTOAIAGLRALAGQAHAPGVNP-A 623  
 QY 535 DSSRKG-LAFLFTGQAGQTLGMRGLYDVWASAPREAFDLCVRLFNOELDRPLREYMAAP 593  
 Db 624 DSGPFGTVFVYSGSGSQWAGMGROLADEPFAAAVAVLEPVEVQAFSLHDVL--- 679  
 QY 594 ASVDAALLDQTAFTQPALETFEYALALMRSGVBEVLVAGHSIGELVACVAGFSLSD 653  
 Db 680 --ANGELVGIQIQTGLTGMQALTELWCSTGYRDLVIGHSWGVAAVAVAGALTPE 737  
 QY 654 AVFLVAAGRLMQALPAGG--AMVSIAPBEADVAAVAPHAASVSIANVANPDQVYIGA 711  
 Db 728 GLRVATSRMLAPLSGGQGMALLELDATTEALADFP--QVTLGINSRQVYINAP 794  
 QY 712 GQPVHAIAMAAARGARTALHVSHPSPMLAPMLAEAFVAESVYSRPSIVLVSNS 771  
 Db 795 TEQIDELIARVRAQNRFASRVNI EVAPHPNADALQPARSELADLPRTPTIGIISTY 854  
 QY 772 GKACTDEVSSPGVWRHAREVVRFADGVKALHAAGG-----TPREVCP-----KSTLLG 821  
 Db 855 ADLHTQPVFADHMAWTNMENPVRFQ--QAIASAGSGADGAVHTLEISAPHLTQALID 911  
 QY 822 LVPACMPDARPALILASSRAGDEPATVLEALGLMAVGLVSMAGLFP-----SGGRVP 876  
 Db 912 TLHSAQPGARYSLSGLQRTDQDVTFRTNLKAHTI-----HPHTHPREPHRP 962  
 QY 877 LFTYVQREYWIIDTKVADDAARGRAPAG-----HDEVEEGAVAGGDRRSARL--DH 929  
 Db 963 IFTTWMQHRHMITTKYPAGSVGS--APRAGTLLGQHTVAVTASPSHLMQARLAPDA 1020  
 QY 930 PPEEGRR---EKTEAA-----GDRPFLELDERV 957  
 Db 1021 KPYQGHREHVEVVPASVVLHTLILSATELGYALSSEVRFPQIFADPRILQV---V 1076  
 QY 958 LDHLVLVTERA-----P 971  
 Db 1077 ADNRAISLASSPACTPDRMTRHTYTAQSSPSDSASSLHNHRANQOPPRARDLIP 1136  
 QY 972 GLEVEIADAAGLSF-----NDVQALGVNPDLLGKPNPPLLGLGECAGRIV 1020  
 Db 1137 DLAEILMRKIDGLPFSWTVASMTQHSNLTVAIDL--PEALPEGSTGPLL--DAAVILA 1192  
 QY 1021 AVGEGVNG-LVVGQVIALSAGAFHTVTSAL-----VLPRQALSIEAA 1068  
 Db 1193 ALSDVADSRLLVPASIEQISLGDVVVTGPRSVTLNKRTAHDDGITVDVTVAHAGEVPSLS 1252

QY 1069 M-PVAVLTAWYALDRIRALQGERVYLHAATGVG-----LAAVQW-----AQHV 1112  
 Db 1253 MRSIALRYALDPLDGLD-VGAQP-----PASTGPYEAVCDAITNFVHTIDMQGTVDATHP 1305  
 QY 1113 GAE--VH-----ATAGTPEKRAYLESIGVRYVSRSRSPFVAD----- 1148  
 Db 1306 GAEQVTHGPVAILIGDDGALCETLEAGQYOPAVMSDQSAARYVYVADSDPAGADETD 1365  
 QY 1149 ---VRAVTGEGS----- 1157  
 Db 1366 VDFAVRITCTETGLVRLTAEADADKPAALMLITRGVHSAVPSALROSFLMGLAGVIAAE 1425  
 QY 1158 ---VDVUNLSLSELIDKSFNLRSRGFVELKRD----- 1190  
 Db 1426 HPELWGLVLDALINDDGEFGPALAEILAKPSKI-LVRDQVVLAPALAPYRGPARKS 1484  
 QY 1191 --CYAD-----NQLGL-----RPFRLNLSFSLVD----- 1212  
 Db 1485 LQCRPDAAVYLITGGIGALGLMAADWLDRGAHRLVLTGRTPLPERRDQDLDTLRLRR 1544  
 QY 1223 ---LRGMMLERPAR-----VRALL-----EELGLI----- 1235  
 Db 1545 IDALRALEMKGVTYEAVAADVGCREDVQALLAARDGGAAPIRGIIHAAGITNDQLVISM 1604  
 QY 1236 ---AAGVFTPP-----IATLP 1249  
 Db 1605 TGDVAVQVWPKIGSSQVYLHDAFPPGSVDFFYLTASAGIGFIPQGSYAANSYLDALA 1664  
 QY 1250 IARVADARSM-----AQOHLGKLYLTLDPE----- 1277  
 Db 1665 RARROGCHTWSLDMVAMRGILAAADQVSEELARMSRITSEAEFTAFEFVDGYOVA 1724  
 QY 1278 --VOIRIPTHAGSGSTDRDLRLBSAAPAR--AALBAFLRTQVSQVLRTEIKV 1333  
 Db 1725 QAVVMPAPRAGADSGANAYLLPRKMSVMAATEVNSELEGLRRIITAEIRVPEKED 1784  
 QY 1334 AEAFLTRIGMDSLMAVELRNRIEASLKLKSTFLSTSPNIALAONLLDALATLSER 1393  
 Db 1785 TDRPRAELGLNSLWMAIRREABQVGELEATVLMFHPPTYKSLASYL-----AKR 1835  
 QY 1394 VAENLRAGVQNDVYSSGA 1412  
 Db 1836 VAPHDVQODNQISALSSSA 1854  
 RESULT 8  
 WA EMENT  
 ID WA EMENT STANDARD; PRT; 1986 AA.  
 AC Q03149;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Confidial green pigment synthase (EC 2.3.1.-).  
 GN WA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC Eurotiales; Trichocomaceae; Emericella.  
 OX NCBI\_Taxid=162425;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RX MEDLINE=93101122; PubMed=1465094;  
 RA Mayorga M. E.; Timberlake W. E.;  
 RT "The developmentally regulated Aspergillus nidulans wa gene encodes a  
 RL polypeptide homologous to polyketide and fatty acid synthases.";  
 RL Mol. Gen. Genet. 235:205-212(1992).  
 CC -1- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN  
 CC INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED  
 CC BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE  
 CC ASSEXUAL SPORES (CONIDIA).  
 CC -1- COTRANSLATION: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES  
 CC (POTENTIAL).  
 CC -1- PATHWAY: Confidial green pigment biosynthesis.  
 CC -1- SIMILARITY: Contains 2 acyl carrier domains.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X65866; CAA46695.1; -.
CC PIR; S28353; S28353.
CC InterPro; IPR001227; Ac transferase.
CC InterPro; IPR000794; Ketoacyl-synt.
CC InterPro; IPR006163; Pp bind.
CC InterPro; IPR006162; pAntenn. attach.
CC InterPro; IPR001031; Thioesterase.
CC Pfam; PF00698; Acyl_transf; 1.
CC Pfam; PF00109; ketoacyl-synt; 1.
CC Pfam; PF02801; ketoacyl-synt C; 1.
CC Pfam; PF00550; pp-binding; 2.
CC Pfam; PF00975; Thioesterase; 1.
CC PROSITE; PS00012; PHOSPHOPANTHETINE; 1.
CC PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
CC TRANSFERASE; P550075; ACP DOMAIN; 2.
CC TRANSFERASE; Phosphopantetheine; Multifunctional enzyme; Repeat.
CC FT DOMAIN 529 582 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC FT DOMAIN 991 1024 ACYL/MALONYL TRANSFERASES
CC (BY SIMILARITY).
CC FT DOMAIN 1650 1719 ACYL CARRIER (ACP) 1.
CC FT DOMAIN 1772 1841 ACYL CARRIER (ACP) 2.
CC FT ACT_SITE 548 548 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC FT ACT_SITE 1001 1001 ACYL/MALONYL TRANSFERASES (BY
CC SIMILARITY).
CC FT BINDING 1682 1682 PHOSPHOPANTHETINE (BY SIMILARITY).
CC FT BINDING 1804 1804 PHOSPHOPANTHETINE (BY SIMILARITY).
CC FT BINDING 1804 1804 PHOSPHOPANTHETINE (BY SIMILARITY).
CC SQ SEQUENCE 1986 AA; 216634 MW; 74EF0940FF40EE9A CRC64;
CC -----
Query Match 16.4%; Score 1180; DB 1; Length 1986;
Best Local Similarity 26.9%; Pred. No. 1.6e-54;
Matches 419; Conservative 242; Mismatches 635; Indels 264; Gaps 53;
CC -----
CC 8 RAADPPIAVGASCRLEPGVILDSGFMTLEGRDVTGRVPAERMDAAAFDPDPDPAAGK 67
CC 373 RAELSKAIITIGMSGRFP-EADSPQDPFNLLYKGLDVHVKYEDERWDADAHVDTLGTATNT 431
CC 68 TPVTRASFSDVACFDASFFGISPREALRMDPPARLLLEVCWELEMAALAPSLVTE- 126
CC 432 SKVPYGCWIREPGLFDRFFNMSPREALQADPAQRLALTYAELLEGAGVDPSTPSTOR 491
CC 127 --TGVFIGIPSEYEALPQATASAEIDAHGGLGTMPSSVGASRISYALGRPCVAADTA 184
CC 492 DRGCIFFGMSDDRR---EYNSGQDIDYFIFPGGNAPFTPGRIYTFKXSGBSVSDTA 547
CC 185 YSSSLVAHLACOSLRSCECTALAGSVLSLSPSTLVLSTKTRALARDRCFAFSAAD 244
CC 548 CSSLIAIHLCNSIMRNDCTAITGGVNIITPNDAHGLDRGHFLSRGTGNCNTPDDGAD 607
CC 245 GFEGGECATVVVLRSLGAAADGRILAVTRGSLINNDGASSGLTVNGSSQELVKRAL 304
CC 608 GYCADADVGVTLRLSDALADNDPILGVINGATYNSAEAVSITRPHVGAQAFIFKLL 667
CC 305 ADAGCAASVGVYEAHGTGTLGPIEIQALNAVYGG--RDVATPPLIGSVKNLGAPE 362
CC 668 NEANVDKRNISYIMHGTCQAGDAVEMQSVLVFAADHRGPGQSLHGSASNSIGHGE 727
CC 363 YASGITLLKRVLSLQHGQIPAHLLAQ--LNPRIWMDLRLVTTRATPMDWNT- 417
CC 728 SASGVTLVKKVLLMKMKMIPPHCGIKTKINHPPTDLAQNNVHIALQ- TANNRPSEFGK 786
CC 418 RRAGVSPGMSGTAAHVLBEAPATCTPPAPF---RPAELVLKARTASAL---DAQ 469
CC 787 RQIFLNNSAAGNTALLLEDGPVSD---PEGEDKRRTHVITLSARSQIALONNIDAL 841

```

```

QY 470 AARLDHLETV---PSQICGVAFSLATTTSRAMEHRLVAATSRBGLRAALDAAAGQ-- 524
QY 842 CQYISQEQETFEVKQSNMLPSLAYTTTARINHPRVVAIGSSPQEMROSLIASSRKEV 901
QY 525 ---TSPGAVRSIADSSRGKLAFLTTGCAQTLGMRGGLYDWSAFREFDL--CYRLFN 578
QY 902 AVPAKPPG-----IGFLFTGQGAQVAAAGKOLYEDCSFRSRIEHLDCISQ-G 948
QY 579 QELDRPLREVMMAEPASVDALLDQAFQPLFFEEVLAALMMSWGEPLVAGHSIG 638
QY 949 QDPLSTPLVDGSLPLSEISPVVQLGTT-----CYVALSSFFMASLGTTFSPVLGHSIG 1003
QY 639 ELVAACVAGVFLSDAVFLVAARGIM--QALPAGAWVSIAPADVAVAAPPAASVS 696
QY 1004 DPAANMAAGVLSSTDITVACGRADLTFRCCPGTHAMLAIAPLVEVQLL--NEKYND 1061
QY 697 IAAVNAPODVITAGAGQPVHAIAMAMARGATKALHSHAHSPMLAMLEAFGRVAES 756
QY 1062 MACINPSESTVISGPSSSIDELSRACSEKGLSTLITVYAFHSAQVEPILDEKALQG 1121
QY 757 VGRPRPSIVLVNLSGACTD---EYSSPGYVVRHAREVVRPADGVKALHAAGAG---T 809
QY 1122 ITFNKSPFVVALGVEITTEGSSNLNLEYLVHRCRETVNLSAFEAARNAKLGGDTLL 1181
QY 810 FVEVGKSTLLGLVPACMPDARPALIASSRAGRDEPATVLEALGIMAVGLVSWAGL-- 867
QY 1182 MLEVGHTYCSGMYATL--GPQTTMASLRBEDTWMKVLNSLSLTYLAGVDIINWQYHQ 1240
QY 868 -FPSSGRVPLPTYPQORRWID-----TKADDAANGDRARAGAGHDEVEGAYNG 919
QY 1241 DRESSHVRVPLPYKWKDLKNYWI PYRNNECLTKGSSMS---AASASLOPTEFLTSAQGV 1296
QY 920 GGRSARL-----DHPPESGR---REKYEAGDRPFRLIDEPPVLDLVLRTER 968
QY 1297 VESRDGLTATVYVANDIADPDLNRYIQCHKNGALCSSLYADSAQTLAEYLI---EK 1353
QY 969 RAPGLGEVEIAVDAAGLSPNDVQLAGVDPDLPGKPNPPLLGEACAGRIVAEGCVNG 1028
QY 1354 YKREL-----KGSGLDVCNTV-----PKRLIAKTKGEGRISATANWVDK 1394
QY 1029 LVVGQVYALSAG-----AFATHVTTSAALVPRPALSAITMAAMP--- 1070
QY 1395 HVSQVQFVSATAGKLLIDAHGEVKLFDCAADLEWKRSYLVK-RSIELLENSAVKGA 1453
QY 1071 -----VAYLTAMVALDIRARLQPERVLI-----HAATGVGGLAAVQ-----WAQ-- 1110
QY 1454 HRLRGVTKLFSALVDYDENTQSTREVLIDSEHETATLVKFOAPQANFHNPIYIDSF 1513
QY 1111 -HV-GAEVHATAGTPEKRAYLESIGRVYVSDRS--DRFVADV-----RAWT 1153
QY 1514 GHLSGFINMASDQTSK-----SQVFVNHGMDSMRCLKFSADVTYRTYVRMQPWRDSIWA 1569
QY 1154 GG---EGVDVY-----LNSLSGELIDKSF-----NLARSHG--KFVELGRKD 1190
QY 1570 GNVYIFEGDIIIAVFGVRFQALSRKIIDIALPPAGLSAQTSPQSSAPQKTEFAKPT 1629
QY 1191 CYADNOLGRPFLNLSFSLVDLRCWMLERPARVRLLEBELGLLIAAG-----VFTPPPI 1245
QY 1630 SRPAPVYTKSFYKKSAGSVVYRL-----NILASVGLSSDSMDLVPFDYGV 1680
QY 1246 ATLPIARVADAFRSMQAQHLGKLYTLTGDPVEYQIRIPT-----HAGAGPST 1292
QY 1681 DSLSLTLVIGKRE-----ELNDM--DSVFIEHPVTGDFKRFVTOQLSPSVASDSSS 1731
QY 1293 GDR-----DLIDBLAS--AAPAAALAEFLRQVQVLR-----TPREIKVGAEL 1337
QY 1732 TDRESEYFRNGDSCSGLSPPASPGTVSPNEKTOIHEMGTKMEIRAIITADEIGVASADI 1791
QY 1338 -----FTRLGMSIIMAVELRNRIEASIKLKTSTFLSTSPNIALLAQNLLDALATLSPE 1392
QY 1792 KSDENINELGMSLSTLVLGKIRSLDMDL-----PGEFFINQTLDOETLADLDK 1843

```

RESULT 9  
MSAS\_PENPA STANDARD; PRT; 1774 AA.  
AC P22367;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 6-methylsalicylic acid synthase (EC 2.3.1.165) (6-MSAS).  
OS *Penicillium patulum* (*Penicillium griseofulvum*).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; microsporitic Trichocomaceae; *Penicillium*.  
OC NCBI\_TaxID=5078;  
RN (1)  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RC STRAIN=DSM 62862;  
RX MEDLINE=91006137; PubMed=209605;  
RA Beck J., Ripka S., Stegner A., Schlitz E., Schweizer E.;  
RT "The multifunctional 6-methylsalicylic acid synthase gene of  
RT *Penicillium patulum*. Its gene structure relative to that of other  
RT polyketide synthases.";  
RL Eur. J. Biochem. 192:487-498(1990).  
CC -1- FUNCTION: This multifunctional enzyme is a polyketide synthase.  
CC It catalyzes a total of 11 steps by seven different component  
CC enzymes, in the biosynthesis of the antibiotic patulin.  
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH = 6-  
CC methylsalicylate + 4 CoA + 3 CO(2) + NADP(+).  
CC -1- PATHWAY: Patulin biosynthesis.  
CC -1- SUBUNIT: HOMOMULTIMER.  
CC -1- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.  
CC -1- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT  
CC FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND  
CC THIOLEASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL X55776; CAA39295.1; -.  
DR PIR: S13178; S13178.  
DR InterPro: IPR001227; AC transferase.  
DR InterPro: IPR000794; Ketoacyl-synt.  
DR InterPro: IPR006163; Pp-bind.  
DR InterPro: IPR006162; Ppantenn-attach.  
DR Pfam: PF00698; Acyl\_transf\_1.  
DR Pfam: PF02801; ketoacyl-synt\_C\_1.  
DR Pfam: PF00550; pp-binding\_1.  
DR PROSITE: PS00012; PHOSPHORANTHENEINE, 1.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE, 1.  
DR PROSITE: PS00075; ACP DOMAIN; 1.  
DR Multifunctional enzyme; Transferase; Antibiotic biosynthesis; NADP;  
KW Phosphopantetheine.  
FT DOMAIN 186 238 ACYLTRANSFERASE (AT).  
FT FT 642 676 ACETYLY/MALONYL TRANSFERASES.  
FT FT 1403 1450 2-OXOACYL REDUCTASE.  
FT FT 1700 1769 ACYL CARRIER (ACP).  
FT NP\_BIND 1419 1424 NADP (POTENTIAL).  
FT ACT\_SITE 204 204 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
FT ACT\_SITE 653 653 MALONYLTRANSFERASE (BY SIMILARITY).  
FT BINDING 1732 1732 PHOSPHOPANTHENEINE (BY SIMILARITY).  
SQ SEQUENCE 1774 AA; 190232 MW; 05EDDD10863F938 CRC64;  
Query Match 16.3%; Score 1177; DB 1; Length 1774;  
Best Local Similarity 25.6%; Pred. No. 2e-54;  
Matches 458; Conservative 210; Mismatches 655; Indels 466; Gaps 56;

QY 74 SELSDVACFDASFFGISPREALRMDPAHRLLEVCMEALENNAAPASALVGTETGVIGI 133  
DB 95 YFLDLRLEDPDOQFFGISPKBAEQMDPQQRVSLSEVASEALBDGIIAKSLSGSDTAVFQV 154  
QY 134 GPSEYEA---LPQATSAEIDAHGIGTTPSVAGAGISYALGRGCVAVDTYSSSL 189  
DB 155 NSDDYSKVLVDLP-----VWEAMWGIGITVYCGVPRNISYHLNLMGPRSTADDAACASLL 208  
QY 130 VAVHLACOSLRSGECSTLAGVSLMLSPSTLVMLSKTRALARDBRCFAFAEADQFGRG 249  
DB 209 VAIHGVQAIRLGESEKVAIVGVGNALCGPGLTRVLDKAGALISSDSCSCSFDDAHGIVARG 268  
QY 250 EGCNAVVLKRLSGARADDRILAVIRGSAINHDSAGSLTPVNGSSOEIVLKRALADAGC 309  
DB 269 BGAGALVLKSLRHALLDHNDVLAIVKGSAAVOCDGKTGNIMAPNSVAQQLAANNALSAANI 328  
QY 310 AASSGVYEAHGTGTTLDPIEIOALNVAVYGLGRDVATPPLIGSVKTNLGHPEVASGITG 369  
DB 329 DPHTVRYEAAHATSTPLDPTETISAIASVYGDADRPADDPVCYIGSIKPINIGHLEAGAGVWG 388  
QY 370 LKXVLSLQHQIIPAHLAQALNPISMGDLRLTVTRARTPMDNTPRRAGVSSFGMSG 429  
DB 389 FLKAVLAIQKGLPPQANLTKINSRIDKTAGVKVQATPWPESDPIRRAGVCSYGYG 448  
QY 430 TNAHVLEE-APATCTP---PAPERPAELVLISARTASALDAQAARLDHLETY-PSQC 484  
DB 449 TVSHAVIEFEPFILOPDLNGCAVSGFG-LLLSSPOEKRLALQKTRIDMTATGAKXOHN 507  
QY 485 LGDVAFSLATTRSAMEHRLAVALATSRREGRLAALDAAQG--QTSFGAVRSIADSSRGKLA 542  
DB 508 LSDIITLTALTRDRHDHVDYRAALVDDYRDABEQVLAQSLANGVHTFTQGRVLSGDISKDV 567  
QY 543 FLFTQGGQITLGMGGLVDWASAFREAFDLCVRLFNGLDRLPREVMAEPASVDAALLD 602  
DB 568 WFFSGHQWPMQKQL-----IHNPVFPAI-----QPLDLIAEGLSPTEILR 614  
QY 603 QTAFA-----TOPALTFEYALALALMRSGVPELVAGHSIGELVACVAGVSELDVAFL 657  
DB 615 TQDFSSSRVQVILTYVMQIGLSALQSGNITPQAVIGHSVGEIAASVAVAGALSPABGALI 674  
QY 658 VAARGLMQALPAGAMVSIAPREADVAANAAPHASVIAAVNAPDOVINAGAOQPVHA 717  
DB 675 VTRRALLYRQWKGKGMILVLPSEATEEIIIGSRSDLV-VAIDSSPSCCVAVGDDELVAE 733  
QY 718 IAAAMAARGATKALHVAHSHAFSPMLAPMLAEFGV-AESVYRRPSIVLVNSLGSXACT 776  
DB 734 TAEALKAQGVKTFVYKSDIAFHSPPTLNGLVDPRLDVLAEITLSPVSPVKLYSTALADPRG 793  
QY 777 DEVSSPGVWVRHAREVRFADGVKALHAAAGATFVEVGPKSTLLGLVPACMPDAR----- 831  
DB 794 QDLRVEVWAGMVMNRVLTSAVKAAYVEDGRLFEVSTHPVSHSINETLMDAGMEDFA 853  
QY 832 --PALLASSRAGDEPAT--VLEALGLMAVGVSWAGLFPSSGRRVPLPTYPQOREY 887  
DB 854 VLPFLT-----KKPTEKHILHSIAQHCAGAEVMAAQN-GRWATGVPTTWMHKRI 906  
QY 888 WIDTAADAAGDRAP--GAGHVEVEGG-----AVRGDR--RSARLDH--PPE 933  
DB 907 W-----RIETAPLHTGLTHD-VEKHLLAGRIIPVPGDITVYTRTLDNDTKPFG 956  
QY 934 S---GRREKVEAG-----DRPFRLEI----- 952  
DB 957 SHPLHGTIVPAAGLINFGLKTGQGMQNVVLRVPVAINAPRSQVVVVQDDQVAVVRL 1016  
QY 953 --DEPGVLDHLVLRVT-----ERRAPGLGE-----VEIAV 980  
DB 1017 IPSEBSQDDDDASWYTHTTAYWDRKVASSEDRIDPAVKSRLVTXLADNFSIDYLGKV 1076  
QY 981 DAAGLSF-----NDVOLALGM-----VPDDL- -GK- -RNP 1007  
DB 1077 SAMGPFMAVVTBYHRYDKEMLARVDVNPALISGADPLPMDSSWAPVLAATAVSSTIFFPT 1136

```

QY 1008 PLLL-----GGECAIRIVAVG----- 1023
DB 1137 ALAMPQIERVEVFTSDPPKISMLVQGEASDVPSHVSVEAGEVLAFTAMRPER 1196
QY 1024 EGVNGL-----VVGQPVIALSAGAFATHTTSAALVPPQALIS-AIEAAMP----- 1070
DB 1197 EGPVGSQSMESLVHQ--IAMPATPABEPISIEITVILVSPDATTALVAASIPTRVNSF 1254
QY 1071 -----VAYLL-----AMVLDJIALDQPE-- 1090
DB 1255 QFSSTQEFFSNASSLPLEKGTVTYITIGEVASLAEPVASESFTWLLLEIKFTVNSGLP 1314
QY 1091 -RYLIHAATGGVGLAAVQMAQ-----HYGAEVHATAGT-----PEKRAYLESL 1132
DB 1315 IKFTILTANIGEGOTPTALAQSPLYGLARIYASHPDLGLIDVEEVIPLSTMRITYQGA 1374
QY 1133 GVRVYSD--SRSDRFVADV-----AMTGEGEV----- 1158
DB 1375 DIRINDGIARTSRFRSLPRNKLPLASEGPRLLPREGTYLITGGGLVGLLEVADPLVEK 1434
QY 1159 -----DVVNSLSGELIDKSFNILRSHGRFV-----ELGKRDQYAD 1194
DB 1435 GARRLLIISRRLPPRRTWQVSEDLQPTI--AKIRLSESGASVHVLPLDITKPPDAVEQ 1492
QY 1195 NQGLRPFRLNLSF-----SLVDLRG-----NMLEPAPARVALLLEELGLIAAGVFTPP 1243
DB 1493 -----LTTALDRLSIPVQGVVHAAGVLDNELVMQTTDAENRVAPARIAGALALHEVFP 1548
QY 1244 P-----IATLPI--ARVADAF----- 1257
DB 1549 KSVDFEVMFSSCGNLVGTQASGSGNAFLDLTATRRALGLDAVASFQWTSRGLDMGA 1608
QY 1258 -----RSMQAQHLGL-----VLTGDPRE 1277
DB 1609 STDFINAELESKGTITVTRDEAFRAWQHLLAKYMDHGVILRSRAFEDEGPIPVSLINDIA 1668
QY 1278 VQIRIPTHAGAPSTGDRDLRLSLASAPARAALAEFLRTQVSOVLRTPEIKVGAEL 1337
DB 1669 VR-RVGVTSNTSPRAAGSS--DAVPTSGPELK-AYLDEKIRGCYAKVLQMTAEVDSKAA 1724
QY 1338 FTRIGDMSLMAVELNRNRIEASLKIKISTFTPLSTSPNIALIALDLALA 1386
DB 1725 LADLVGDSVMTVTLLRQLQTLKIAVPPTLWHSPTSLHAWFAEKLA 1773

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: I42766; AAC41675.1; -
DR EMBL: I42766; AAC41674.1; -
DR PIR: T17490; T17490.
DR InterPro: IPR001227; Ac. transferase.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR006163; PP. bind.
DR InterPro: IPR006162; Phantne. attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl-transf. 1.
DR Pfam: PF02801; ketoacyl-synt. C. 1.
DR Pfam: PF00550; pp-binding. 1.
DR Pfam: PF00975; Thioesterase; 1.
DR PROSITE: PS00075; ACP_DOMAIN; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; FALSE NEG.
KM transferase; Acyltransferase; Phosphopantetheine;
KM Multifunctional enzyme.
FT DOMAIN 374 805 BETA-KETOACYL SYNTHASE.
FT DOMAIN 1714 1785 ACTL_CARRIER (ACP).
FT DOMAIN 543 543 THIOESTERASE.
FT ACT_SITE 993 993 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 993 993 ACTL/MALONYL TRANSFERASES (BY SIMILARITY).
FT BINDING 1746 1746 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT_SITE 1937 1937 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 2109 AA; 230715 MW; CB701372A16D8551 CRC64;

Query Match 16.3%; Score 1173.5; DB 1; Length 2109;
Best Local Similarity 27.2%; Pred. No. 3.8e-54;
Matches 409; Conservative 222; Mismatches 625; Indels 247; Gaps 54;

QY 9 AADPPIATVAGACRLPGVYIDLSGFWTLLGSRDITVGRPAERWDAANFPDPDAPGCT 68
DB 369 AGCKXALAIYMSGRFPESPTTES-FWDLLYKGLDVCKEYPRRRWDINTHVDPGKARNKG 427
QY 69 PVTBASFLSDVACFDASFPGISPREALRMDPAHRLLEVCWEALNAALASALVGTG- 126
DB 428 ATKMGCMIDBSGDFDRFGIISPKXAPQMDPROMALNTEYEMARAGLVPTTSTQKD 487
QY 127 -TGVFTIGPSEYEALPQATASAEIDAH---GLGTPSVGAGRISYALGRCVAV 181
DB 488 RIGVFHGVTSNDM---METNTAQNDITYFITGNNGFIP---GRINCFEFGASVYTN 539
QY 182 DTVASSSLVAVHLACOSLSGECSTLACGVSLMSPSTLWLSKTRALARDGRCAFS 241
DB 540 DTRCSSSLAIHILACNSLWKGCDITVAGGTWNIYTPDGHGTGIDKGFSLRTGNCCKPYD 599
QY 242 EAGFPRGSCAVVYKRLSGARADGRIIAVIRGSAIHNHDASSGLTVPNGSSQEIYLV 301
DB 600 KADGYCRABGVGTVPFKRLLEDALADNDPLVGLDKTNHSAMSESMTRPHVGAQIDNMT 659
QY 302 RALADGCAASVGVYEAHGTGTLGDPLEIQALNAVYG---LGRDVATPLLGIVKTNL 358
DB 660 AALNTTGLHNDSYIEHMGITGVQVDVAVMESVLPAPSEFARADQPLFYGSAAKAV 719
QY 359 GHPEVAGTGLKLVVLSLOHGOIPHL--HAQALNPR1-SWGDRLVTVTRAKTPWPDW 414
DB 720 GHSEGVSQVSLIKVLMQMHDITIPPHCGIKPSKINRNFPDYGABNVHIAFEKWPRT 779
QY 415 NTEPRAGVSSFGMSGTNAHVLEAPATCTPPAPRPAPLIVLSARTASALDAQARLR 474
DB 780 HTRRVVLINFSAAAGGTALIVDAPBRHMPTEKDRSSHIVALSAAVGSAMKTNLERLH 839
QY 475 DHELYPSQCLGDVAESLATTRSAMERHLAVALTSREGRLAALDAQAQ---GOTSQAVR 531

```



```

Db      840  QYLLKNPTDIAQLSYTTTARMHYLVHVSVTGAEVEERTKLENAIONGDVSRP----- 895
Qy      532  SIADSSRKLAFLFTGQAGQTLGMGRGLYDVWASREAFDLCVRL-----FNQELDR 583
Db      886  ----KSKKILFAFGGGSQYATMKQYDAIPSRREDLEKDRLAQSHGPFSPFHVCTS 951
Qy      584  PLREVMMAEPASVDAALLDQTAFTOPALFTFEYALAAWMRSWGVEPELVAGHSIGELVAA 643
Db      952  PKGDVEEMAPVV-----QLATTCQMALTINMTSGIRPDVTGHSIGEFALL 1000
Qy      644  CVAGCFSEDAVFLVLAAGRLMQLPAG--AMVSIAPENDVAAVAAPHAASVIAAVN 701
Db      1001  YAAAGVLSADVVYLVGORAELLGRCORGTAMLVKATPEALSOIMDH--DCEVACIN 1058
Qy      702  APDQVVIAGQOPVIAIAAAMARGARTKALVSHAFSPLEMAPLLEAFGRABESVSR 761
Db      1059  GPEDVVLGGTTKNVAEVOBAMNDNGIKCTLLKLPFAFSAQVQPLDDPEALAGATPAK 1118
Qy      762  PSIVLVSNLSGKACTDE--VSSPGVWRHAREVVRPADGVKALHAAG---AGTFEVGPX 816
Db      1119  FOLLISPLRTEIHHEGGVTPSYVAQHCRHTVDMAQLRSAREKGLIDDKTLVIELGPX 1178
Qy      817  STLGLVPCMPDAPAL--LASSRAGDEPATVLEALGGLMAVGGLVSWA---GLPFGS 871
Db      1179  PLISGMVMTVLGDKISTLPTLAPNKA---IWPSTOKILTSYVTGMDINMKYHAPFAS 1235
Qy      872  GRVPLPYPMQREBYWIDTKAD---DAARGRRAPAGHDEVEGAVRGDRRSALD 928
Db      1236  QKVVDLPFYGMDLKQYIPIYQGDWCLHRHQDCKAPRHE-----IKTADYQ----- 1283
Qy      929  HPPESGREKVEAAGDRPFRLIDE---PGVLDELVL--RVTERRAPGLGEVEIYVDAAG 984
Db      1284  -VPEPS-----TPHRPSKLDPEKEAPEIKTTTLHRVEETTKLGN--LVETD 1332
Qy      985  LSFNDVQ-LALGWPDLPKKNPPLLIIGBECAGIYAVGE-GVNGLVGQPVIALSAGA 1042
Db      1333  ISRKDVNGLARGHLVDGI-----PLCTPSFADIAMQGVQYSQRLRAHP---GAGA 1382
Qy      1043  FATHTTTSALV---LPR---POAL-SAIEAAMPVYLTWVYLDLIAR-LQGEENVL 1093
Db      1383  IDGLVDSDMVVDKALVPHKGPPOLLRTTLTWMPKAAATRSKVKVFATYFADGKLD 1442
Qy      1094  IHA-----TGGVGLAAV-----QWAOHVGAEVHATAGTPEKRAVLESIGVYVSDSR 1142
Db      1443  EHASTVAFSTDQOLKSLRSVSEKTHI-RQLHGHAKGQCMRYNRTGYKXLMASMARF 1501
Qy      1143  ----DRFVADVRAWTGEGVDVINSLSGEL-----I 1170
Db      1502  NPDVWLLDYLVLTNEAENEAAGVDPSLSSSEGTFAAHPAHVDAITOVAGFAMNANDV 1561
Qy      1171  DK-----SFNLL-----RSHGRVEVLGKRCVADNOL-----GLRPLR 1204
Db      1562  EKQYVNVHGMDSFQYQPLDNSKSYQVYTKMQA---KENDLVHGDVVVLDEQVAFPR 1618
Qy      1205  NLSFSLVDLRGMLERPVRVALLLELLGLIAGVFT---PPPIATPLIARVADAFRSM 1260
Db      1619  GLTLRSVVRGALRVVLQTTVKADQL-----GKTMPSPPPTTTPISGYKRAANOV 1672
Qy      1261  ----AOQHLGKVLTLGDEVOIRIPTAGAGSPSTGDRDLRLASAAPARAAA-- 1312
Db      1673  SSOAIPAEATH-----SHTPPQ-----PKISPV-PET-----AGSAPAAAGVGS 1711
Qy      1313  ---LEAFRTQVSOVLRTPEIKVGAEL-----FTRLGMSDLMAYELANRIEASIKLUS 1364
Db      1712  NEKLDAVVRV-VSE-----ESGIALLEBLTDSNPFADMGIDSLSSVIGSRFREDGLDLG 1765
Qy      1365  TTF 1367
Db      1766  PEF 1768

```

RESULT 11

```

FAS_RAT
ID_FAS_RAT STANDARD; PRT; 2505 AA.
AC PI2785; 009187; 009190; 064717;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN FASN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89240686; PubMed=2717611;
RA Amy C.M., Witkowski A., Naggert J., Williams B., Randhawa Z.,
RA Smith S.;
RT "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT fatty acid synthase."
RL Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Liver;
RC MEDLINE=93075999; PubMed=1339331;
RA Beck K.F., Schreglmann R., Stachopoulos I., Klein H., Hoch J.,
RA Schweizer M.;
RT "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT norvegicus."
RL DNA Seq. 2:359-386(1992).
RN [3]
RP SEQUENCE OF 75-2505 FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RC MEDLINE=89128431; PubMed=2915923;
RA Schweizer M., Takabeayashi K., Beck K.F., Schreglmann R.;
RT "Rat mammary gland fatty acid synthase: localization of the
RT constituent domains and two functional polyadenylation/termination
RT signals in the cDNA."
RL Nucleic Acids Res. 17:567-586(1989).
RN [4]
RP SEQUENCE OF 2085-2505 FROM N.A.
RX TISSUE=Mammary gland;
RC MEDLINE=88087240; PubMed=2891707;
RA Naggert J., Witkowski A., Mikkelsen J., Smith S.;
RT "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT domain of the rat fatty acid synthetase."
RL J. Biol. Chem. 263:1146-1150(1988).
RN [5]
RP SEQUENCE OF 1921-2324 FROM N.A.
RX TISSUE=Mammary gland;
RC MEDLINE=87246646; PubMed=3109907;
RA Witkowski A., Naggert J., Mikkelsen J., Smith S.;
RT "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT protein and its flanking domains in the mammalian fatty acid
RT synthetase."
RL Eur. J. Biochem. 165:601-606(1987).
CC - FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC ACYL CARRIER PROTEIN.
CC - CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+)
CC - CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC acetyl-[acyl-carrier protein].
CC - CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC - CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC [acyl-carrier protein].
CC - CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC - CATALYTIC ACTIVITY: (3R)-3-hydroxyacylmalonyl-[acyl-carrier protein]
CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.

```

CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-  
 CC 2,3-dehydroacyl-[acyl-carrier protein] + NADPH;  
 CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-  
 CC carrier protein] + oleate.  
 CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: M76767; AAA57219.1; -  
 DR EMBL: X62886; CAA44679.1; -  
 DR EMBL: X62889; CAA44680.1; -  
 DR EMBL: X13415; CAA31780.1; -  
 DR EMBL: X13527; CAA31882.1; -  
 DR EMBL: J03514; AAA41144.1; -  
 DR PIR: A30313; XYRTPA.  
 DR InterPro: IPR001227; Ac transferase.  
 DR InterPro: IPR002085; Adh\_zn family.  
 DR InterPro: IPR000794; ketoacyl-synt.  
 DR InterPro: IPR006163; Pp bind.  
 DR InterPro: IPR006162; Pantane attach.  
 DR InterPro: IPR000051; SAM bind.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam: PF00698; Acyl\_transf. 1.  
 DR Pfam: PF00107; ADH\_zinc\_N. 1.  
 DR Pfam: PF02801; ketoacyl-synt\_C. 1.  
 DR Pfam: PF00550; pp-binding. 1.  
 DR Pfam: PF00975; Thioesterase. 1.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.  
 DR PROSITE: PS00606; B KETOACYL SYNTHASE; 1.  
 DR PROSITE: PS50075; ACP DOMAIN; 1.  
 DR Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;  
 KM Hydrolyase; Oxidoreductase; Transferase; Lyase; NADP;  
 KM Pyridoxal phosphate.  
 FT DOMAIN 1 413 BETA-KETOACYL SYNTHASE.  
 FT DOMAIN 429 817 ACYL AND MALONYL TRANSFERASES.  
 FT DOMAIN 1629 1857 ENOYL REDUCTASE.  
 FT DOMAIN 1858 2113 BETA-KETOACYL REDUCTASE.  
 FT DOMAIN 2118 2174 ACYL CARRIER (ACP).  
 FT DOMAIN 2202 2505 THIOESTERASE.  
 FT ACT\_SITE 161 161 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
 FT ACT\_SITE 581 581 MALONYLTRANSFERASE (BY SIMILARITY).  
 FT ACT\_SITE 1665 1682 NADP (ER).  
 FT BINDING 1698 1698 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT NP\_BIND 1765 1780 NADP (KR).  
 FT BINDING 2151 2151 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT ACT\_SITE 2302 2302 THIOESTERASE (BY SIMILARITY).  
 FT ACT\_SITE 2475 2475 THIOESTERASE (BY SIMILARITY).  
 FT ACT\_SITE 878 878 BETA-HYDROXYACYL DEHYDRATASE (BY  
 FT ACT\_SITE SIMILARITY).  
 FT CONFLICT 871 871 S -> P (IN REF. 3).  
 FT CONFLICT 1967 1968 MV -> IL (IN REF. 5).  
 FT CONFLICT 2085 2085 C -> P (IN REF. 4).  
 FT CONFLICT 2106 2106 A -> V (IN REF. 1 AND 5).  
 FT CONFLICT 2296 2296 Y -> H (IN REF. 1 AND 5).  
 SQ SEQUENCE 2505 AA; 272647 MW; 5810EC13D37F3114 CRC64;

Query Match 15.38; Score 1100; DB 1; Length 2505;  
 Best Local Similarity 22.78; Pred. No. 3.4e-50;  
 Matches 470; Conservative 221; Mismatches 586; Indels 792; Gaps 56;

QY 12 DPALVAGSGLRPGVIDLSGFMTLLGSRDTCVRVPAERDAAMFDPDPDAFGKTPVT 71  
 DB 2 EEVVIAMSGKLPESE-NLOEFMANLIGVDVW-TDDRRWKAGLYGLP-----K 49  
 QY 72 RASFSLVACFDASFFGISPREALRMDPARLLLEVCWEALENAIAPSAIVGTETGVFI 131  
 DB 72 RASFSLVACFDASFFGISPREALRMDPARLLLEVCWEALENAIAPSAIVGTETGVFI 131

DB 50 RSGKLKLSKDFDASFFGVNPKOAHMTMDPOLRLLEVESEYAVIDGGINPASLRGTNVGVV 109  
 QY 132 GIGPSEYEALQOATASAEIDHAGLGTMPSGAGRISVALDRPCVAVDPAVSSIVA 191  
 DB 110 GVSGSEASBALRDEPT-LIGYSVWGCORAMNARLSFFDFKPSIALDPAICSSSLA 167  
 QY 192 VHLACOSLRSGCESTALAGVSLMSPSTLWMLSTKTRALARDGRCAFEAGDFGRGCG 251  
 DB 168 LONAYOATISGSCPAIVGVIGILLKPTNSVQPMKLGMSLPQTCRSFDSGNGYCRATA 227  
 QY 252 CAVVYLKRLSGARADGDRILAVIRGSAINHGA-SGLTVPNGSGOEIVKRALADAGCA 310  
 DB 228 VVAVLLTKKSLAR---RYATILNAGVTIDCKEKGVTFFPSGEAQEQLIRLSLYPGGVA 283  
 QY 311 ASVGVVEAGTGCTTGDIETIOALNAVYGLGRDVA---TELLIGSVTNGHPHYASG 366  
 DB 284 PESLEYIEAHGTVTKVDP--OELN--GITRSLCAFPQSPSLIGSTIKSNNGHPBPAG 337  
 QY 367 ITGLKVLVSLQHGQIPALHQAQALNPR-SWGDRLRTVTRARTPMDNTPRR--AGV 422  
 DB 338 LAALTIVLSTLNGVAPRLHFNHNPBEIPALDGRLOVV-----DRPLVRGIGVGI 390  
 QY 423 SSFGMSGTNAHYVL---EAPRATCTPPAPERP-AELLVSAKTSALDAQAARLDL 477  
 DB 391 NSFEGFGANVHILQNTQOAPA---PAPHAALPHLHAAGRTMEAVQGLEQGRQHS 445  
 QY 478 ETPPSQCLGDVA.PSLATTSAMEHRLAVATSRREGURLALDAAGQTSPPAVRSIADS 537  
 DB 446 Q-----DLAF-----VSLMNDIAATPTAMPRTGVTVLQVESHVO--EVQVPAHQ 489  
 QY 538 RGLAFLFTGGAQTLGMGRGLVDVWASAPREAFDLCVRLFNQELDRPLREVMABPASYD 597  
 DB 490 R-PLWFTICGMGTQRMGLSLMRL-DSFRESI-----LRSDALKPLVKKVSDLLSD 542  
 QY 598 AALIDQTAQPALFTFEYALALMWSGVBPVLVAGHSIGELVAAVCVAFSLDAVYL 657  
 DB 543 EHTFDIDIVSFSILTAIQIALIDLLTSMGLKDKDGIIGSLGAVACGYADGCLSORAVLA 602  
 QY 658 VAARGLMQ--ALPAGNAVSIAPRADVAAVAAPPAVSI.AAVNAPQVVIAGAGQVY 715  
 DB 603 AWRGGCICIDANIPA-GSMAAVGLSWECKQCRP--GVPRCHNSEDTVTIISGQAAV 658  
 QY 716 HAIAAAMARCAARTKALVHS-AFHSPLM---AP-MLEAFGRV-----AESVYRRP 762  
 DB 659 NEVEEDKQEGVPAKEVTRTGLAFHSYFMEGIALPTLQALKKYIRPRRSARWLSTJSP 718  
 QY 763 STLVSNLSKCACTDEVSSPGVVRHARREVPADGVKALHAAGCTFVEVGPSTL--- 819  
 DB 719 EAQWQSSLA-----RTSSAEYVNVNVLVSPVLFOEAL--WHVPEHAVVLEIAFHALLQAV 770  
 QY 820 --LGVPAQMPDARPALLASRAGRDEPATVLEAGGLMAVGGLVWAGLFPSCGRVP- 876  
 DB 771 LKRGVPRSC---TIPLMKDKDNLEFPLTNLGAHVLGTIDIPNALFFPVEPVR 825  
 QY 877 ----LPTVPWQERYWIDTKADDAAG-----DRRA- 903  
 DB 826 GTPFLISPHIKMHSQWMDIPVADFPNGSSSSATGVYNIDASSESDHYLVHDCIDGRVL 885  
 QY 904 -PGAGH-----D 909  
 DB 886 FPGTGLVYLWVKTLASLSLEETPVVENVTFHQATILPRTGTVPLEVRLLEASHAPE 945  
 QY 910 EYVEEGAVRGG-----DRSARLDHP--PPES----- 934  
 DB 946 VDSQNLIVSGKYYQWEDPDSKLFDPBVPPIPAESESVSRLTQGEVYKELRGLGYGRH 1005  
 QY 935 ----- 934  
 DB 1006 FQGVNATLEGBQCKLMDKNWTFMIDTMLQISILGFSKQSLQLPTRVTALYIDPATHLQ 1065  
 QY 935 ----- 934  
 DB 1066 KYVMEGDTQVADVTTSCIGVTVSGGVYISRLQTTATSRQOEOQLVPTLEKEFVFTPHVE 1125

```

QY 935 -----GREYEAACGRP----- 947
Db 1126 PEGLSAATLQKELQCKGLAKALQTKATQOGLKMTVPQJEDLPQNGLPRLLAACQQL 1185
QY 948 ---FRLEIDEP----- 955
Db 1186 NGNLQLELGEVLARELLPEBPLISGLNSQALKACIDTALENLSTLKKVVEVLAGEG 1245
QY 956 -----GTL 958
Db 1246 HLYSHSALLNTQPMQLQLEYTATDRHPQALKQVQTKLQOHDAVAGQWDSGAPPTNLGAL 1305
QY 959 D-----HLVLR----- 964
Db 1306 DLVNCALATLGDPAALADNMVAALKDGGFLMHTVLGSHALGETLACLPEVQGPSPF 1365
QY 965 -----VTERRA----- 970
Db 1366 LSQEWESLPSRKALHLVGLKKSFGYATLFLCRSLSPQDKPIFLPVEDTSFGWVDSLKSI 1425
QY 971 -----PGLGE 975
Db 1426 LATSSQPVMLTAMNCPISGVVGLVNCLEKPEGRHRCILLSNLSSTSHVPKLDPGSSE 1485
QY 976 VEIAYDA----- 982
Db 1486 LQKVLSEDLVMNVYRDGANGARHHPQLEDDKEBEQTAHAFVNVLTGDLASIRWSSPLK 1545
QY 983 -----AGLSFNDVOLALG-NVPDDLPGK-PNPPLLLGEGCAGRIYAVG 1023
Db 1546 HMQPPSSGSAQLCTVYVYASLNFEDIMLATGKLSFDPAIPGKMASRDCMLGMEFSGR-DKCG 1604
QY 1024 BEVNGLVGQPIYALASAGAFATHVTTSALVLPRLPALSAREANPVAVLTAMVLDLI 1083
Db 1605 KRWGVLV-----PAEGLATSVLLSPDLMDVPSSWTLERASVPVYTTAYSLVYR 1656
QY 1084 ARLOPGERVLIHAATGCVGLAAVQMAQHAEGVAEYHAGTPEKRAYLES-----LGVRVYSD 1139
Db 1657 GRIQHGELYLIHSGSGCVQMAISLISGCRVFTTVSAGAKRAYIQAFPPQLDDTSPAN 1716
QY 1140 SFSDFRADVRAWTGGEGVDVVLNSIGSELIDKSFNLLSHGRFVGLGRDYADNQLG 1199
Db 1717 SBDTSPEQVHLHTGGKGVDLVNSLAEBKLQASVACLAHQHGFLEIGKFDLSNNHPLGM 1776
QY 1200 REFPLRLTSYLDKRMMLERAPVALLBELLGLIAAGVTFPPATIPARVADAFS 1259
Db 1777 AIFLNKVTFHGLLDLLEFGANDSWREVAELLKAGIRDVVVKLCTVFPKQVEDAFRY 1836
QY 1260 MAOQHGLKVLTLGDPVEQIRIPTHAGAGPSTGDRDLRLASAPARAALAEFLRT 1319
Db 1837 MAQGHKIGVIVQVREPEEMLP--GAQPT-----LISLSKTFCEHKSYYIT 1884
QY 1320 -----OVSQ-VLRTP-----IKVGAEL-----FTRLGMSLMAVELNRIE 1356
Db 1885 GGLGFGELRLARWLVRGAQRLVLTSGSIRGCVQAKHREWRQGIHVLTSTNVSSLE 1944
QY 1357 ASLKLKLTSTFTSTFNIALLAQNLLDAL 1385
Db 1945 GARALIAEATKLGPGVGFNLAMVLRDAM 1973

```

```

RESULT 12
STCA EMENT
ID STCA EMENT STANDARD; PRT; 2181 AA.
AC Q12397;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Putative sterigmatoctystin biosynthesis polyketide synthase (PKS).
GN STCA OR PKST.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

```

```

OC Eurotiaceae; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 26;
RX MEDLINE=96202293; PubMed=8643646;
RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA Keller N.P., Adams T.H., Leonard T.J.;
RT "Twenty-five conserved transcripts define a sterigmatoctystin gene
RT cluster in Aspergillus nidulans";
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=95370159; PubMed=7642507;
RA Yu J.-H., Leonard T.J.;
RT "Sterigmatoctystin biosynthesis in Aspergillus nidulans requires a
RT novel type I polyketide synthase";
RL J. Bacteriol. 177:4792-4800 (1995).
CC -1- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
CC STERIGMATOCTYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC -1- CORRECTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETHEINES
CC (POTENTIAL).
CC -1- PATHWAY: Sterigmatoctystin biosynthesis; first step.
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: U34740; AAC9191.1; -
DR EMBL: U39121; AA81586.1; -
DR HSSP: P25715; 1MLA.
DR InterPro: IPR001227; Ac transferase.
DR InterPro: IPR000794; Ketoacyl-synth.
DR InterPro: IPR001633; pp_bind.
DR InterPro: IPR001632; Pantine_attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl_transf. 1.
DR Pfam: PF00109; ketoacyl-synth. 1.
DR Pfam: PF02801; ketoacyl-synth. C. 1.
DR Pfam: PF00550; pp-binding; 2.
DR Pfam: PF00975; Thioesterase; 1.
DR PROSITE: PS00075; ACP DOMAIN; 2.
DR TRANSFERASE: ACYLTRANSFERASE; Phosphopantetheine; Repeat;
KW Multifunctional enzyme.
FT DOMAIN 383 814
FT DOMAIN 884 1209
FT DOMAIN 1706 1777
FT DOMAIN 1830 1901
FT ACT_SITE 2 2181
FT ACT_SITE 552 552
FT ACT_SITE 978 978
FT BINDING 1738 1738
FT BINDING 1862 1862
FT ACT_SITE 2028 2028
FT ACT_SITE 2181 AA; 238831 MW; 5A3B5712AA9AD942 CRC64;
SQ
Query Match 15.0%; Score 1083; DB 1; Length 2181;
Best Local Similarity 25.9%; Pred. No. 2.3e-49;
Matches 402; Conservative 235; Mismatches 637; Indels 280; Gaps 53;

```

```

QY 7 EBAADP-----IIVGASGRLPGGVLDLGGFTLLSGSRPTVGRPAERDAAMFDP 60
Db 370 EYSHRPGSDRGKALIVMSGHFPAP-STDSFMVLTKGLDYCKEVPARRMDVTVTHDP 428
QY 61 PDABGKTPVTRASFSLDVACPDASFQGISPREALRMDPAHRLLEVCMALENNAIAPS 120

```

Db 429 SGKARKKATRCWGCWIDFAGEFDPFRFSSISPKKAPQWDPQRMALMSTYEMERGGIVDP 488  
 Qy 121 ALVGTG---TGVTIGIGPSEYEALPQATASAEIDH---GGLTGMPSGAGRISALG 173  
 Db 489 TTPSTORNRIGVHGVTNDW---METNTAONIDVTYITGKRGFP---GRINCFEE 540  
 Qy 174 LRPCCAVDPAVSSSLVAVHLACOSLRSGECSTALAGVSLMSPSTLVWLSKTRMLARD 233  
 Db 541 FSPSPSNDTACSSSLAHLACNSLMKRGCDTAVAGTMMIPTPDGHTLDDKGFSLKRT 600  
 Qy 234 GRCKAFSAEADGFRGEGCAVVVLKRLSGADGDRILAVIRGSALINHDGASSGLTVPNG 293  
 Db 601 GNCKAFDDADGYCRAEGVGTVPFKRLLEDALAEENDPLATILDIKTHSMSMSMTRPFK 660  
 Qy 294 SSGEIVLKRLLAAGCAASVGVVEAHGTTGTDGPTEIOALAAVYG---LGDVATPRL 350  
 Db 661 PAQIDNMSSALLSTAGISPLDLSTYIEMHGTOGVDAEMESVLSLFAPDFTFPRDKPLY 720  
 Qy 351 IGVKTNLGHPEYASGITGLKTVLSLHGQIPAHLLAQAALNP--RI--SMGDL---RLT 403  
 Db 721 VGSAGKANIGHGEVSGVTSLSIKVLLMKNDTIPEHC---GKFGSRINRNYPPPLPARNVH 777  
 Qy 404 VTRAKTPMDMNTPRRAGVSSFCMSGTNAHVLEAPAACTCPPEPAPPELLVLSART 463  
 Db 778 IAFEPKPMPTDTPRVLINNFSAGGNTAVLVEDAPV-----DELGEIAGSFG 826  
 Qy 464 SAlDAQAARLDLETPSQCLDVAFSLATTTSSAMHRLAV--AATSRBELRALDLAAQ 522  
 Db 827 EAAGNOSSEL-----SYTTARRMHHPRVSTIGANTMELLRVESIAR 871  
 Qy 523 GQSPCAVRSIADSRGKLAFLFTGOGAQTLCMGRLGYDVWSAFR---EAFDLCVRLFNQ 579  
 Db 872 GH---GYNR---PATRKVIYIACSGQSOYTGGMQOYNYPFRSLERFDOLAISYGF 925  
 Qy 580 ELDRPLREVMABPA--SVDAALLDQTAFTQPAIFTEFVLAALMWSGVPELVAGHSI 637  
 Db 926 P---SELEVTYSKRPVGDSEMDL---PVIQALVLSLEMLGMLGSGFKLPKPAVIGHSL 979  
 Qy 638 GELVACVAGVPSLEDVAVELVARGRLMQLPRAG--AMTSIAPEDVAAVAPHASV 695  
 Db 980 GEYALAYISGVLSAADTLVYVGMKAKLQRCORGTAMLAVRASPTLCEVLA--ESNC 1037  
 Qy 696 SIAAVNAPDOVVIAGAGOPVHAIAAAMAARGAARTKALVHSHAFSPMLAEMLEAFGVAE 755  
 Db 1038 EVACHNGENPTVLSGLPKETMMIIONSLSATGICKTLKLFATFHSAGVQIILEFKNVAR 1097  
 Qy 756 SVSYRRPSIVLVNSGKACTDE--VSSPGYVWRHAREVRA---DGVKALHAAAGCTFV 811  
 Db 1098 GVTFFHKQIPLSLPLVAVKVIDEKGTVPVYLARRCHREPRVKNVSVLEHARDQHIITDRIV 1157  
 Qy 812 -EVRGPKSTLLGLVPACPDARPALMLASSRAGRDEPAVLEALGLMAVGVLSWAGL--- 867  
 Db 1158 IDVGPALKMGMKMTLLDKDTSSALPTLGPLSDVWKSLLTILGLYSRGIDIMVVAHER 1217  
 Qy 868 FPGGGRVPLPTYPWQRERYWIDTKADDAAGDBRABGAGHDEVEBGAVRGGRSARL 927  
 Db 1218 FGSAKKVTELPSPYGMWDKQVFIPIKGEWCHL-----RHERGCAIPKGETATSQY 1268  
 Qy 928 DHPPEPSGRREKVEAAGDRPRLFEIDPG---VLHLVLVYTERRAVGLGEVEIAVDA 983  
 Db 1269 QLSDE-----QVAAKRPSKODESKEAVPEIATTVHARVVEKTEPIGAT--LVET 1319  
 Qy 984 GLSPNDV-QIALGVNPPDLP-----GKPNPILLGGCAGRIYAVGSG-VNGLV-VGQPV 1035  
 Db 1320 DISRPDNOIAQGLVNDGIPCLTPSYVADIALHGVRSNMRLRSHDAGGVVADVMV 1379  
 Qy 1036 I---ALSAGAFAHTVTSALVLPPOALSATEAAMPVAVLTWVYLDRIARLQPERV 1092  
 Db 1380 IDKALIHGKSPQRLKRTLLTMTP--PKAAATTSRAKKFATYPRADGLD----- 1427  
 Qy 1093 LHAATGTVGLAAVQMAQHVGAEVHATA-----GTPEKRAYLESIG-----VR 1135

Db 1428 -----TEHATCTVFTSEAOIKSLQKVPXEYQERIKKLGELGRCQCFIR 1471  
 Qy 1136 YVSD-----SRSDRPVADVRAMTGWGEVDVJNSLSELID-KSFNLLRSHGRFVELGK 1188  
 Db 1472 YTKSGYKLMSSMASFHRDYKILN-----HLLINEADNEAVSTMPSAAKSECTF--AA 1523  
 Qy 1189 RDCYAD--NQLG-----LRPFLRNLSPSLV-----DLR 1214  
 Db 1524 HPAVYDAIIVQGVGFAMNANDNTDIOQEVFVNHGWTSSFQYQPLVKGKTYEAVYRWTEDEK 1583  
 Qy 1215 GMLTERPARY---RALLEELGL-IAAGVFTPPPIATLEIAVADAFRMAQAHLGKLV 1270  
 Db 1584 GDLVHDDTIVLYGDAAVAFPKGLSVLSHLSSTQLTSTVYVRVPRRGRLMVLOQASDKA 1643  
 Qy 1271 LTLGDE--VOIRPHTAGAPSTGDRDLDRLASAAPARAALAEAFRTQVSQVLRTEP 1329  
 Db 1644 RLHGNOQAVKTOAFORA-----LKQRPQSSPTPHASKVAYSASISPTAKGV 1693  
 Qy 1330 I-----KYGAELPTRLGMDSIMAVELRNRIEASIK 1360  
 Db 1694 VVARDLREGDDKFKAVLSVISESGVALGELTADTNFADIGDLSISWIGSRLELDIG 1753  
 Qy 1361 LKLSL-----TFLSTSPNIALAQNLLDALATLALSLERVAENL 1399  
 Db 1754 LEIGAFSPSLFDICPTVRSLKTLISGS--AVSVNDKDELDPGQEAETAAPQOL 1804  
 RESULT 13  
 PAS\_HUMAN  
 ID FAS\_HUMAN STANDARD; PRT; 2504 AA.  
 AC P49327;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].  
 GN FASN OR FAS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=96004605; PubMed=7567999;  
 RA Jayaramar A., Tai M.-H., Huang W.-Y., Al-Feel W., Hsu M.,  
 RA Abu-Elheiga L., Chitrals S.S., Wakil S.J.,  
 RT "Human fatty acid synthase: properties and molecular cloning."  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:8695-8699(1995).  
 RN [2]  
 RP SEQUENCE OF 753-758 AND 1285-1297.  
 RX MEDLINE=94294385; PubMed=8022791;  
 RA Khatjida F.P., Jemner K., Wood F.D., Hennigar R.A., Jacobs L.B.,  
 RA Dick J.D., Pasternack G.R.,  
 RT "Fatty acid synthesis: a potential selective target for  
 RT antineoplastic therapy."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6379-6383(1994).  
 CC - FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF  
 CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.  
 CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN  
 CC ACYL CARRIER PROTEIN.  
 CC CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a  
 CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).  
 CC - CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +  
 CC acetyl-[acyl-carrier protein].  
 CC - CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +  
 CC malonyl-[acyl-carrier protein].  
 CC - CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-  
 CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
 CC [acyl-carrier protein].  
 CC - CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +  
 CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
 CC - CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein]

```

CC      = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.
CC      -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC      2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC      -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl]-
CC      carrier protein + oleate.
CC      -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC      -1- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND
CC      LIVER.
CC      -1- MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY
CC      MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTHETHEINE CONTENT OF
CC      THE PROTEIN.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U26644; AAC50259.1; -.
CC      PIR: A57788; A57788.
CC      GENE: HENC:3594; FASN.
CC      MIM: 600212; -.
CC      InterPro: IPR001227; AC transferase.
CC      InterPro: IPR002085; Adh zn family.
CC      InterPro: IPR000794; Ketoacyl-synt.
CC      InterPro: IPR006163; pp_bind.
CC      InterPro: IPR006162; ppantne_attach.
CC      InterPro: IPR001051; SAM bind.
CC      InterPro: IPR001031; Thioesterase.
CC      Pfam: PF00698; Acyl transferase.
CC      Pfam: PF00107; ADH zinc N: 1.
CC      Pfam: PF00109; ketoacyl-synt; 1.
CC      Pfam: PF02801; ketoacyl-synt_C1.
CC      Pfam: PF00550; pp-binding; 1.
CC      Pfam: PF00975; Thioesterase; 1.
CC      ProSite: PS00606; B KETOACYL SYNTHASE, 1.
CC      ProSite: PS00075; ACP DOMAIN; 1.
CC      ProSite: PS00012; PHOSPHOPANTHETHEINE; FALSE NEG.
CC      Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
CC      Hydroxase; Oxidoreductase; Transferase; Lyase; NADP;
CC      Pyridoxal phosphate.
CC      DOMAIN 1 413
CC      FT DOMAIN 428 815 BETA-KETOACYL SYNTHASE.
CC      FT FT 1650 1857 ACYL AND MALONYL TRANSFERASES.
CC      FT DOMAIN 1658 2113 ENOYL REDUCTASE.
CC      FT FT 2118 2174 BETA-KETOACYL REDUCTASE.
CC      FT DOMAIN 2202 2504 ACYL CARRIER (ACP).
CC      FT ACT_SITE 161 161 THIOESTERASE.
CC      FT ACT_SITE 580 580 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
CC      FT ACT_SITE 876 876 MALONYLTRANSFERASE (BY SIMILARITY).
CC      FT ACT_SITE 876 876 BETA-HYDROXYACYL DEHYDRATASE (BY
CC      FT SIMILARITY).
CC      FT NP_BIND 1666 1683 NADP (ER).
CC      FT BINDING 1699 1699 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC      FT NP_BIND 1880 1895 NADP (KR).
CC      FT BINDING 2151 2151 PHOSPHOPANTHETHEINE (BY SIMILARITY).
CC      FT ACT_SITE 2302 2302 THIOESTERASE (BY SIMILARITY).
CC      FT ACT_SITE 2475 2475 THIOESTERASE (BY SIMILARITY).
CC      SEQUENCE 2504 AA; 273100 MW; 8AA99809B2338DFA CRC64;
CC
Query Match 14.4%; Score 1039; DB 1; Length 2504;
Best Local Similarity 23.5%; Pred. No. 5.6e-47;
Matches 477; Conservative 208; Mismatches 582; Indels 766; Gaps 68;
OY 12 DPALVAGSCLPGVIDLSGFWTLLESRTYGVKVPKPERMDAAMPDPDDAGKPTVT 71
DB 2 EEVYIAGFGKLPESE-NLOEFWMDLIGVDMV-TDDRRKRYAGLYGLP-----R 49
OY 72 RASFSLVACDPAFFGISPEALAMDPAHRLLEVCWEALENNAIAPSLVGTGVFI 131
DB 50 RSGKTKLDSRRDASFVGHVPRQAHNMDQLNLLLEATYEAIVDGINDSLRGHTIGVW 109

```

```

Db 1137 OLCKGLVEALETKVQOGLKMWVVDWGTGPRSPRPSQGEIPRLLSAACLRLQNLQLEL 1196
      886 -----RYMTDKADD-----AARGD--RRARG 905
Db 1197 AQVLAQERPKLPEDPLLGLLSDPALKACLDTAENNPSSLMKVVEVLACHGLYSRI PG 1256
      906 -----AGHDEVEEGAVRG----- 919
Db 1257 LLSFPHLLQSYATATDHPQALEAAQAELOQHVAQCMPPADAPASLSADLLVNCNA 1316
      920 ----GDRRSARLD-----HP-----PPESGR----- 936
Db 1317 VAALGPASALSNMVAALREGFLLTLRGHPDRIVAFLTSTPEQYGGILSQDAMES 1376
      937 ----REKVEAG-----DRPRLEID----- 953
Db 1377 LFSRVSRLVLGKSPFYGATFLCRPPTQDSPIFLVPTDSTFRWESLKGILADEDSR 1436
      954 -----EPG-----VLDH 960
Db 1437 PVLKAINCATSGVGLVNCLEPBGTVACVLLSNSTSHVEVDPGSAELQKVLQCD 1496
      961 LVLRVTERRAPGL-----GE 975
Db 1497 LNMNVVRDGMGVFRHFLLEDKPEPTAHAFVSTLTGRDLSIRWCSLRHAQPTCPGA 1556
      976 VELAVDAAGLSFNDVOLALG-MWDDLRGK-PNPRLLLGEGCGRIVAVEGNGVLVQ 1033
      1557 QLCVTYVASINFRDIMATKSLSPDALPGKTSQDSLSIMFSGR-DASGRVWGLV-- 1612
      1034 PVIALSGAPATHTTSAALVLRPPQALSAIEAAMPVATLTAWALDRJARLOPGERVL 1093
      1613 -----PAKGLATSVLSPLDMVPSNWTLEBAASVAVYSTAYALVAGRRPGETTL 1667
      1094 IHAATGCVGLAAVQWAGVAGVATAGTEPKARVLES-----LGVVSDSRSDRPVADY 1149
      1668 IHSGSGGVQAAIALALSTLCRFVTTGSAEKAAYLQARPPQDSTSFANSRPTSFQHV 1727
      1150 RAMTGGCGVUVVNSLSELIDKSFNLRSHGRFVEIGKDCVADNOLGRLPRLNLSFS 1209
      1728 LMTGGGVDLVNLSAEELQASVRCFTHGRLELTKGDKLSONHPLGMAITLKNVTFH 1787
      1210 LVDLRGMLERPARVRALTEELGLIAGVFPPTPIATLPIARVADAFRSMAOQHLGL 1269
      1788 GVLDAFNNSSADMREVVALVEAIRDGVVRPKCVFHGAYQEDBFRTMAGGKHIGV 1847
      1270 VLTGDBEVOIRIPTHAGAGPSTGDRDLRLASA-----APAR-----AAALEAFLRTQ 1320
      1848 VVQVLAEEPAVL-----KGAKP-----KLMSAISKTFCPAHKSYIIAGGLGCF-GLE 1893
      1321 VSGVLRPEIKVGAE--ALFTPLGM-DSLMAVELRNRLEASLKLKJSTTFLST 1370
      1894 LAQWL-----IQRGVQKVLTSRSGIRTGQAKQVRRRQGLQOVSTSISS 1942

```

## RESULT 14

```

PKSK_BACSU STANDARD; PRT; 4447 AA.
AC P40803;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksk (PKS).
OS PKSK.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / PB1424;
RX MEDLINE=95219083; PubMed=7704258;
  Albertini A.M., Caramori T., Scoffone F., Scotti C., Galizzi A.;

```

```

RT "Sequence around the 159 degree region of the Bacillus subtilis
RL genome: the pkex locus spans 33.6 kb.";
RT Microbiology 141:299-309(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Aevero V., Besterio L., Bessieres P., Bolojin A., Borchert S.,
RA Bortles R., Bourster L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.U., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Deniot F., Devine K.M., Dunethof A., Ehrlich S.D., Emerson P.T.,
RA Eutian K.D., Errington J., Fabre C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Hega K., Hitech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogawa K., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Puig P., Putrelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha R., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekouchi J., Sekowska A., Serot S.J., Serot P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler H., Weitzenger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yatsunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
  subtilis."
RL Nature 390:249-256(1997).
CC - FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
  THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
  SECONDARY METABOLISM.
CC - COFACTORS: CONTRAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
  (POTENTIAL).
CC - SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
  FAMILY.
CC - SIMILARITY: Contains 5 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.tsb-sib.ch/announce/
  or send an email to license@tsb-sib.ch).
CC EMBL; U11039; AA85144.1; -
CC EMBL; 299112; CAB13590.1; -
CC EMBL; 299113; CAB13601.1; -
CC PIR; A69679; A69679.
CC HSSP; P14687; IAMU.
CC Subtilat; BG10930; pksk.
CC InterPro; IPR000873; AMP-bind.
CC InterPro; IPR001242; Condensaatn.
CC InterPro; IPR000794; Ketoacyl-synt.
CC InterPro; IPR006163; Pp_bind.
CC InterPro; IPR006162; Pplantn_attach.
CC Pfam; PF00501; AMP-binding; 1.
CC Pfam; PF00668; Condensation; 1.
CC Pfam; PF00109; ketoacyl-synt; 3.
CC Pfam; PF02801; ketoacyl-synt_C; 3.
CC Pfam; PF00550; Pp-binding; 5.
CC PRINTS; PR00154; AMPBINDING.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.

```

DR PROSITE; P500455; AMP BINDING; 1.  
 DR PROSITE; P500606; B KETOACYL SYNTHASE; 2.  
 DR PROSITE; P50075; ACP DOMAIN; 5.  
 KM Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;  
 KM Phosphopantetheine; Multifunctional enzyme; Repeat; Ligase;  
 KM Complete proteome.  
 FT DOMAIN 1 68 ACYL CARRIER (ACP) 1.  
 FT DOMAIN 1063 1130 ACYL CARRIER (ACP) 2.  
 FT DOMAIN 2516 2589 ACYL CARRIER (ACP) 3.  
 FT DOMAIN 2618 2687 ACYL CARRIER (ACP) 4.  
 FT DOMAIN 3868 3937 ACYL CARRIER (ACP) 5.  
 FT BINDING 31 31 PHOSPHOPANTHETHEINE (POTENTIAL).  
 FT BINDING 1093 1093 PHOSPHOPANTHETHEINE (POTENTIAL).  
 FT BINDING 2552 2552 PHOSPHOPANTHETHEINE (POTENTIAL).  
 FT BINDING 2650 2650 PHOSPHOPANTHETHEINE (POTENTIAL).  
 FT ACT SITE 2915 2915 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
 FT BINDING 3900 3900 PHOSPHOPANTHETHEINE (POTENTIAL).  
 FT ACT SITE 4147 4147 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
 SQ SEQUENCE 4447 AA; 496058 MW; 9C6DB4A6C0A9C057 CRC64;

Query Match 14.1%; Score 1018; DB 1; Length 4447;  
 Best Local Similarity 26.3%; Pred. No. 1.4e-45;  
 Matches 391; Conservative 203; Mismatches 519; Indels 376; Gaps 52;

QY 2 ADPIERAAE-DPIAIVGASCRLLPGGVIDLSGFWLLBGRDPTVGRVPERMDAAAF-D 59  
 DB 2733 ABRNKKQADFEPAIVGISGRFP-GAMIDEFWKNLEEGKDSITEVPRDRMDREHYGN 2791  
 QY 60 PPPDAGKTPVTRASTLSDVACFDASFFGISPREALRMPARHLLLEVCEALENAIAP 119  
 DB 2792 PPTDV-NKTDIKWGGFIDVAFDPLFFGISPREADYVDPQOQLMTYWKWLEDEGCS 2850  
 QY 120 SALVGETGVFICIGPSEYEALPQATASAEIDAHGGLTM-PSVAGRIYALGLRGC 178  
 DB 2851 QSLSTGTCIFIGTGTGYKDLFHR-ANLPLEGHAATHMLPSVGNMSTFLLNHGS 2908  
 QY 179 VAVDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLVMLSKTRALARDRC 238  
 DB 2909 EBEVETACSSSLVAIHRAVTAQMONGCEMAIAGVNTILTEBAHISYKAGMLSTORCK 2968  
 QY 239 FSAEADGFRGCGCAVYVVKLSGARADDRILAVIRGAINHDCASGLTPVNGSQGT 298  
 DB 2969 FSAADNGVYRGVGVNMLKLEDAERDGNHIVGVRTAEHNGGANTLTSPNPAQAD 3028  
 QY 299 VIKRALADAGCAASSGVYEAHGTGTLGDPTEIQALNAVY---GLGRVAPRL----- 350  
 DB 3028 LTVRAVROADIDPSTVYTIENAGTGTGLDPIEINGLKAFAKELSMRSESQDPVDHRC 3088  
 QY 351 -IGVYKTLGHEBYASGITGLKVLVSLQHQI PAHLHAQALNPRIISMGDLRLTVTRAT 409  
 DB 3089 GIGVYSNIGHLELAAGISGLIKVLLQMGKTLVSLHCETLNPYLQGLDPSFYIQEKO 3148  
 QY 410 PWP-----DMN-TPRRAGVSFGMSGSTNAHVLEE-APATCTPPAPERPAELVLISART 462  
 DB 3149 EKSVTYDRGNELPRAGISSFGIGVNAHIVIEETMPANSEHTTEOP-NVIVLSAKN 3207  
 QY 463 ASALDAQARLDHL--ETYPSCGCVAFSLATTSAMEHLLAVATSRREGIRALDA 520  
 DB 3208 KSRLLDRASQLEVRNKKYTDODLRIRIYTLQVGEEDERACVAGTMOELEKQLRF 3267  
 QY 521 AOGQTSPGAIVRSIADSSRGKLAFLFTGCAQTLGMRGLYDVWSAFREAFDLCVLFNOE 580  
 DB 3268 VDGKETDEFFRGQSHRNKETOTIFPADDMALAL-----DAWIRKRYAKLA----- 3315  
 QY 581 LDRPLREVMWAPASVDALLDQTAFTOPALTF-----EYALALMRSGVPEELV 632  
 DB 3316 -----DLWVGVSIONNTL--YGETKRELISLPSYPAPKDHVWPAHESEERKKELY 3366  
 QY 633 AGHSIGELVAACVAGVSLDEAVFLVAAGRLMOALPAGAVSIEAP-ANVAANA-- 689  
 DB 3367 -NALEDRAACFLTKOMSLSP-----IGSAVPGRTVAILCCQGTADLAEVSSY 3414  
 QY 690 -BHAASVSIAAVNAPO-----VVIAGAG-----QPVHAIAMAAARCA 727

DB 3415 FNNHLLIDVSRLE-NDQSDIDWKEFDGLVDVIGCGMDDEGRDLMIEWOQLVEFGHKEGL 3473  
 QY 728 R-----TKL-----HVSIAFIISPLMAPLLEAFGRVAESVSRPBI 764  
 DB 3474 RLLCVTKGLESPQNTSVMAAGSRAGLYRMQCESSHLSIRMDA-----EEVTDRR--- 3526  
 QY 765 VLVSNLSGKACTDESSPGYWRHAREVVRFPADVKAHLHAAGATFVEVGPKSTLLGLVP 824  
 DB 3527 -----LAKLADFEYSYSYDE-----VCYRDLGR-----YQAFKAPETGATBOS 3569  
 QY 825 ACPDAPRALLASRAGDEPATVLEALGGLVAVGGLVSMAGLFPSCGGRVPLPTYPQOR 884  
 DB 3570 AVFP-----KDH-----VLLIGTRGIGLCA----- 3592  
 QY 885 EYVWIDTKADDAARDRAAPGAGHVEVEGGAVRGDRRSARLDHPPESGR----- 937  
 DB 3593 -RHFAB-----CYVKKLVLTGRBOL-----PPREMARFKTSNTS 3627  
 QY 938 --EKVEAAGDRPFRLEIDEPVL-----DHLVLRTERRAPGLGEVIAVDAAGL 985  
 DB 3628 LAEKIQAV-----RELAKGVQVEMLSLTSDDAQVQGTQHIKRTIGPISGVHICAGL 3681  
 QY 986 SFNDVQALGWPDDLPCKPNPRLLLGECAGRIYAVGEGVGLVVGQPVIALSAGAFAT 1045  
 DB 3682 TMDTLAIFIRKTSDDIQVLEPK-----VSGLT-----TLVR 3713  
 QY 1046 HTTSAALVLRPQALSI--EAAAMPVAYLTAMVALRIARLOGERVLHATGCVGL 1103  
 DB 3714 HVCNPELOFVULFSSVSAIIPELASGADYMANSYMYPFAAHQKAPII----- 3764  
 QY 1104 AAVQNA--OHVGAEVHATAGTEPKRAYLESICVRYVSPSRSPFVADVRAWTGEGVDV 1161  
 DB 3765 -SVQPMNKEYS-----MGEVTHQAYRDS--GLSITNSBGRFIDQVSKKFG--PVY 3813  
 QY 1162 LNSLGEIDIKSFMILSRHSGRFVELGKRCYADNQLGRLPFLRNLSFSLVDRCMMLEP 1221  
 DB 3814 LPAMA-----NQTWNEPEL-----LMKRX 3833  
 QY 1222 ARVRLLEELGLIAGVTPPTPIATLPIARVADFRSMAQOHQKGLVLTGDFEVQIR 1281  
 DB 3834 PHEGLQCE-----ALQSPFA-----RDIIEADEVSK----- 3860  
 QY 1282 IPTHAGPSGTGRDLRLRLASAPARAALAEAFRTQVSOVLTPEIKVGAELFRTL 1341  
 DB 3861 -----CGGLSET-----QSLILDLFTBELAIDEDPEIDGLFODY 3896  
 QY 1342 GMDSLMAVELRNRIEASLKLSTFLSTSPNIALLAQNLDAATALS 1390  
 DB 3897 GVDSTILAQVLRIRKLEALDPEILVEYPTIQRFADWLIGSYERLS 3945

RESULT 15  
 FAS CHICK  
 ID\_FAS CHICK STANDARD; PRT; 2511 AA.  
 AC P12376;  
 DT 01-OCT-1998 (Rel. 12, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.39;  
 DE EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].  
 GN FASN OR FAS.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=90311;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.  
 RC STRAIN=white leghorn; TISSUE=liver;  
 RX MEDLINE=95031085; PubMed=7944406;  
 RA Huang W.-Y., Chitkala S.S., Wakil S.S.;  
 "Amino-terminal blocking group and sequence of the animal fatty acid



RT synthase.";  
 RL Arch. Biochem. Biophys. 314:45-49(1994).  
 RN [2]  
 RP SEQUENCE OF 75-1775 FROM N.A.  
 RX TISSUE=Liver;  
 RC MEDLINE=89282777; PubMed=2734291;  
 RA Holzer K.P., Liu W., Hammes G.G.;  
 RT "Molecular cloning and sequencing of chicken liver fatty acid  
 synthase cDNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4387-4391(1989).  
 RN [3]  
 RP SEQUENCE OF 1568-2512 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=89139426; PubMed=2917973;  
 RA Chirala S.S., Kasturi R., Pazirandeh M., Stoloow D.T., Huang W.-Y.,  
 RT "A novel cDNA extension procedure. Isolation of chicken fatty acid  
 synthase cDNA clones.";  
 RL J. Biol. Chem. 264:3750-3757(1989).  
 RN [4]  
 RP SEQUENCE OF 1752-2512 FROM N.A.  
 RX MEDLINE=8830436; PubMed=2842766;  
 RA Yuan Z., Liu W., Hammes G.G.;  
 RT "Molecular cloning and sequencing of DNA complementary to chicken  
 liver fatty acid synthase mRNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).  
 RN [5]  
 RP SEQUENCE OF 2202-2512 FROM N.A.  
 RX MEDLINE=89088152; PubMed=3207710;  
 RA Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;  
 RT "Characterization of a genomic and cDNA clone coding for the  
 fatty acid synthase domain and 3' noncoding region of the chicken liver  
 fatty acid synthase gene.";  
 RL Biochemistry 27:7778-7785(1988).  
 RN [6]  
 RP SEQUENCE OF 2121-2209.  
 RX MEDLINE=89192401; PubMed=2648999;  
 RA Huang W.-Y., Scoops J.K., Wakil S.J.;  
 RT "Complete amino acid sequence of chicken liver acyl carrier protein  
 derived from the fatty acid synthase.";  
 RL Arch. Biochem. Biophys. 270:92-98(1989).  
 RN [7]  
 RP SEQUENCE OF 2209-2508.  
 RC STRAIN=White leghorn;  
 RX MEDLINE=89088151; PubMed=3207709;  
 RA Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;  
 RT "Complete amino acid sequence of the thioesterase domain of chicken  
 liver fatty acid synthase.";  
 RL Biochemistry 27:7773-7777(1988).  
 RN [8]  
 RP SEQUENCE OF 667-674 AND 1698-1709.  
 RX MEDLINE=89323081; PubMed=2751995;  
 RA Chang S.I., Hammes G.G.;  
 RT "Amino acid sequences of pyridoxal 5'-phosphate binding sites and  
 fluorescence resonance energy transfer in chicken liver fatty acid  
 synthase.";  
 RL Biochemistry 28:3781-3788(1989).  
 CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-  
 CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THIS  
 MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN ACYL  
 CARRIER PROTEIN.  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a  
 long-chain fatty acid + (N+1) COA + N CO(2) + 2N NADP(+).  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = COA +  
 acetyl-[acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = COA +  
 malonyl-[acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-  
 carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
 [acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +  
 NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]  
 = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.

CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-  
 2,3-dehydroacyl-[acyl-carrier protein] + NADPH.  
 CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-  
 carrier protein] + oleate.  
 CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.  
 CC -1- ALTERNATIVE PRODUCTS:  
 Name=2;  
 Event=Alternative splicing; Named isoforms=2;  
 Name=1;  
 IsoId=P12276-1; Sequence=Displayed;  
 Name=1;  
 IsoId=P12276-2; Sequence=VSP\_000149;  
 -1- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM  
 POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U04485; AAB46389.1; -;  
 CC EMBL; J03860; AAA48767.1; -;  
 CC EMBL; J02839; AAA62106.1; ALT\_SEQ.  
 CC PIR; S57248; XYCHPA.  
 CC InterPro: IPR001227; AC transferase.  
 CC InterPro: IPR002085; Adh\_zn\_family.  
 CC InterPro: IPR000794; Ketoacyl-synt.  
 CC InterPro: IPR006162; pp\_bind.  
 CC InterPro: IPR001031; Thioesterase.  
 CC Pfam: PF00698; Acyl\_transf.1.  
 CC Pfam: PF00107; Adh\_Zinc\_N.1.  
 CC Pfam: PF00109; ketoacyl-synt.1.  
 CC Pfam: PF02801; ketoacyl-synt\_C.1.  
 CC Pfam: PF00550; pp-binding.1.  
 CC Pfam: PF00975; Thioesterase.1.  
 CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.  
 CC PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 CC PROSITE; PS50075; ACP\_DOMAIN; 1.  
 CC Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;  
 CC Transferrase; Hydrolase; Oxidoreductase; Lyase; NADP; Acetylation;  
 CC Alternative splicing; Pyridoxal phosphate.  
 CC KW INIT MET  
 CC FT DOMAIN 0 7411  
 CC FT DOMAIN 427 815  
 CC FT DOMAIN 1638 1866  
 CC FT DOMAIN 1867 2119  
 CC FT DOMAIN 2124 2180  
 CC FT DOMAIN 2209 2511  
 CC FT MOD\_RES 1 160  
 CC FT ACT\_SITE 160 160  
 CC FT ACT\_SITE 579 579  
 CC FT ACT\_SITE 877 877  
 CC FT NP\_BIND 1674 1691  
 CC FT BINDING 1707 1707  
 CC FT NP\_BIND 1888 1903  
 CC FT BINDING 2157 2157  
 CC FT ACT\_SITE 2308 2308  
 CC FT ACT\_SITE 2481 2481  
 CC FT VARSPIC 2348 2348  
 CC FT CONFLICT 77 78  
 CC FT CONFLICT 116 116  
 CC FT CONFLICT 675 675  
 CC FT CONFLICT 1169 1169  
 CC FT CONFLICT 1178 1178  
 CC FT CONFLICT 1191 1191  
 CC FT CONFLICT 1198 1198  
 CC FT CONFLICT 1286 1287  
 CC BETA-KETOACYL SYNTHASE.  
 CC ACTL AND MALONYL TRANSFERASES.  
 CC ENOYL REDUCTASE.  
 CC BETA-KETOACYL REDUCTASE.  
 CC ACYL CARRIER (ACP).  
 CC THIOESTERASE.  
 CC ACETYLATION.  
 CC BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
 CC ACTL/MALONYL TRANSFERASES  
 CC (BY SIMILARITY).  
 CC BETA-HYDROXYACYL DEHYDRATASE  
 CC (BY SIMILARITY).  
 CC NADP (ER).  
 CC PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC NADP (KR).  
 CC PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 CC THIOESTERASE (BY SIMILARITY).  
 CC THIOESTERASE (BY SIMILARITY).  
 CC T -> TOCPSSEIF (in isoform 1).  
 CC /FTId=VSP\_000149.  
 CC QI -> PV (IN REF. 2).  
 CC L -> A (IN REF. 2).  
 CC R -> S (IN REF. 2).  
 CC K -> N (IN REF. 2).  
 CC A -> T (IN REF. 2).  
 CC R -> H (IN REF. 2).  
 CC P -> L (IN REF. 2).  
 CC DN -> ND (IN REF. 2).

```

FT CONFLICT 1372 1372 K -> E (IN REF. 2).
FT CONFLICT 1533 1533 C -> Y (IN REF. 2).
FT CONFLICT 1577 1577 M -> R (IN REF. 2).
FT CONFLICT 1685 1696 OAAIAIALSMGC -> ASHCHRLHGLA (IN REF. 2;
AAA48767).
FT CONFLICT 1732 1732 Q -> E (IN REF. 2).
FT CONFLICT 1745 1745 S -> N (IN REF. 2).
SQ SEQUENCE 2511 AA; 274648 MW; 622039DAC8315D3F CRC64;

Query Match 13.8%; Score 997.5; DB 1; Length 2511;
Beet Local Similarity 21.7%; Pred. No. 8.6e-45;
Matches 425; Conservative 186; Mismatches 564; Indels 783; Gaps 43;

QY 12 DEIAIVGASCRPLPGVIDLSGFMTLESGRDTYGRVPAERMDAAAFDDPDPAKGTPTV 71
D 1 EDVVIAGIAGKLPESF-NIQEFEMNLNGVDWYTE-DDRRMKPGIYGLP-----K 48
QY 72 RASFISDVACPDASFPGISPRBALMDPAHRLLEVCWEALBNAAIAPALVGTETGPTI 131
D 49 RNGKLDIKKFPASFPGVHPKQAHMTDPOLRLLEVSYEAILDGGINPTALRGTDGVVW 108
QY 132 GIGSEYEALPOATSAEIDAAGLGTMPSVGAGISVALGRGPCVAVDTAVSSIVA 191
D 109 GASGBALFALSO--DPEELGISMTCORAMLANRISTFYDPTGSLITDTCSSSLMA 166
QY 192 VHLACOSLRGECSTALAGVSLMSPSTLVMLSKTRALARDGCKAFSAEADGFGREG 251
D 167 LEMAKAIRHGGCSALVGVVNIILKPNISVQFMKGMISPDGACKAFVSGNGYCRSA 226
QY 252 CAVVVLKRLSGARADGDRILAVIRGSAINHGD-ASSGLTVPNGSGOEIVLKRALDAGA 310
D 227 VVVVLLTKKSMK---RYVATIVAGSNTDGFKEGVTFPSGEMOOLVGLYRECGIK 282
QY 311 ASSVGVENHAGTTLGDPIEIOAL-NAYYGLGRDVAITPLIGSKYTNIGHREYASGITG 369
D 283 PGDVEVEHAGTGTKGDPOEVNGIVNFCOCERE--PLLIGSTSNNGHBPASGLAA 339
QY 370 LKVVLSLOHQOIPAHILHAQALNPRISW--GDLRLTVTRARTPMDMTTPRAG--VS 423
D 340 LAKVLSLEHGLMAPNLHNDPNPDIPALHDSGLKVVCK-----PTVKGGLVSTN 390
QY 424 SFGMSGTNAHVLEB---APATCTPPAPERPAELVLVSARTASALDAQARLRDILE 478
D 391 SFGFGSGNAHVILRNEKKCOFOETCNLP-----RLVQVCGRTOEAVELLIESRKHG 444
QY 479 TYP-SQCGDVAFLATITRSAMEHRLAVAATSREGIRALDDAAOCOTSPGAVRSIAD-- 535
D 445 CSPLSLSLSDIS--AVPVSSMPYR-----GYTLVGTESDITETIQ 481
QY 536 ---SSRGKLAFLFTGGAQTLGMRGLYDVMSAFREAFDLCVRLFNQELDRPLREVMMAE 592
D 482 QVQAGRGPLMYICSGMGTOMKMGSL-----MKDLDFQSLTRS-DELYKST 528
QY 593 PASVDAALL-----DOTAFTOPALFTEYVALAALMRMGVEPELVASHGISGELVAVCV 645
D 529 GLKVSDDLINADENFTDDTVHAFVGLAALQIADIVLKAAGLPDGLIHSHVSELACGYA 588
QY 646 AGVFSLEDAVFLVAARGRLMO--ALPAGAMVSIAPENDVAAVAAPHAASVIAVANAP 703
D 589 DNSLSHEEVLAAYWRCVKEAKLPPGG-MAAVGLTWBECKORCP--NVVPACHNSE 644
QY 704 DOVVIAGAGQPVHAIAAAMAARGARTKAL-HVSHAFFPLMAFMLAFA----- 750
D 645 DTVVYSGPLDSVSEFTTKLKDGCVPAKEVRAGVAFHSTYMASIAPALLSALKYIPIHK 704
QY 751 GRVAVSVYRPSIVLVNLSGKACTDEVSSPGYVWRHAREVVRPADGVKALHAAAGATF 810
D 705 PRSAWISTISIPESQOSDLARN-----SSAEYHVNNLVNPLYLFEHGLK--HIPENAVV 756
QY 811 VEVGPKSTLLGLVPACMPARPALASSAAGRDEPTVLEALGMLVAVGLVSMAGLPPS 870
D 757 VBIAPHALQALIRRTLKPTCTTLLPMKDKHKNLEFLLTQGIKHLTGINVLGNLFP 816

```

```

QY 871 GGRVPL-----PTYWQREERYWIDTKYADDAARGDRAPGA-----G 907
D 817 VEYVPVGTPLISPYIKMDHQDMVPAKEDFPSSGKSASAVYNIIVSPSPHYLVG 876
QY 908 H----- 908
D 877 HCIDGRVLVPATGYLVLAARLARSLGVMMEQTAVMFEVTTIHOATILPKKGSQLEVRI 936
QY 909 ----- 908
D 937 MPASHFEVSGNGNLAVSGKISLENDALKNFHNQADPOSOANVTASGLMEDVQEL 996
QY 909 ----- 908
D 997 HLRGVNPTFGVLECNSEGSAGKILWNGMVTFLDTLHLILAETGRSLPRTIRS 1056
QY 909 ----- 915
D 1057 VYIDVHLQEOVYOYQDNVEAFDVVVDRLDLSLKGAGVOINGLHASVAPRQOERISPTL 1116
QY 916 -----AVRGDRRSARLDHP 930
D 1117 EKFSFVPIESDCLSSSTQILAHYLECKGLLOKLOAKMALHGVKLVHGLETKGAASGP 1176
QY 931 PPSGRRE-----KVEAGDRPFRLP-----IDEP--GVLHVLVRV----- 965
D 1177 PAQKGLQHILREICGELBNGNPHSELBOIVTQEKHLDODPLNLGLDLSSELKTCLDYAK 1236
QY 966 -----TERRAG- 972
D 1237 ENTTSRHKIYEBALAGSRLFSRVQSIINTPOLDYIATDCTBETLSNETELHDAGI 1296
QY 973 ----- 972
D 1297 SFGQDPSSLGSGNLTNADLAVNCSTSVLGNTAEIISNLAAYKEGGFVLHLTLKEET 1356
QY 973 LGEV----- 976
D 1357 LGEIVSLTSPDLQOKHSFLSOAQWELFSKASLNLVAMKSFSGSVIFLCRQSPAKAP 1416
QY 977 -----EIANDA----- 982
D 1417 ILPLVDTHYKXVDSLKEILDSBQPLMTATNGNSGILGMVNCRLBAGHRIRCVF 1476
QY 983 ----- 982
D 1477 VSNLSPSSTVPATSLSSLEMOKIIERDLWMNVYRDCKKGSFRHLPLQOAOPOELTECAV 1536
QY 983 -----AGLSFNDVOLALG-MVPDDLPGK-P 1005
D 1537 NVLTGSDLSLRWIVSLRHHQTTNPVQOLCKVYVASLNPFMDIMATGKLSPDALPGMW 1596
QY 1006 NPPLLGECAGRIYAVEGVNGLVGQPVIALSAGAPATHYTTSAALVLRPQALSAIE 1065
D 1597 LQOCMLGMEFSGRDLA--GRVWGL-----LPAKGATVDDCDKREFLMEVPEMWTLIEE 1647
QY 1066 AAAMVAATLTAMXALDRLARLOPGRVYLIIHAATGVGLAAYQMAOHGVAEYVATGTPEK 1125
D 1648 AASVPAVATAYALVAVRGKMGKGSVLHSGSGGVGAALAIATLSMGRVAFATGSAK 1707
QY 1126 RAYLES-----LGVRYVSDRSRDPFADYRAVMTGEGVUVVNLNLSGELIDKSFNLRSHG 1181
D 1708 REYLQARFPQDANSFASRSNTTPOQHILRVITNGGSLVNLNLSAEKLGASLRCLAHG 1767
QY 1182 RFVELGKDCYADNOLGLRPFLRLNLSFSLVDLRGMLELPPARVRLLELLGLIAGVFT 1241
D 1768 RFLBEGKFDLNNNSQLGALFLKXVAFHGIILDSIFEEGNOEWEEVSELTKGIDGVVK 1827
QY 1242 PPIATLPIARVADAFRMAOAHGKLVLTLDGSEVO 1279
D 1828 PLRTTVFKEEVEAFAFRMAOGKHIGKVIKIQEEBKO 1865

```

Sat Oct 4 14:14:34 2003

us-10-014-717-2.rsp

Page 24

Search completed: October 2, 2003, 17:28:40  
Job time : 54.2576 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 26.2576 Seconds  
(without alignments)  
2544.975 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210

Sequence: 1 VADRIERAADPEIPAIIVGAS.....GVQNDVSSGADQDWIIAL 1421

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2107.5	29.2	3567	1 ERY2_SACER	Q03132 saccharopol
2	1925.5	26.7	3491	1 ERY1_SACER	Q03131 saccharopol
3	1889	26.2	3319	1 OL56_STRAT	Q07017 streptomyc
4	1875	26.0	3172	1 ERY3_SACER	Q03133 saccharopol
5	1852.5	25.7	2110	1 MCAS_MYCBO	Q02251 mycobacteri
6	1592.5	22.1	1538	1 PPSB_MYCTU	Q10978 mycobacteri
7	1583	22.0	1876	1 PPSA_MYCTU	Q10977 mycobacteri
8	1180	16.4	1986	1 WA_EMENI	Q03149 emeticicila
9	1177	16.3	1774	1 MSAS_PENPA	P22367 penicillium
10	1173.5	16.3	2109	1 PKSI_ASPPA	Q12053 aspergillus
11	1100	15.3	2505	1 FAS_FAT	P12785 ratusus norv
12	1083	15.0	2181	1 STCA_EMENI	Q12397 emeticicila
13	1039	14.4	2504	1 FAS_HUMAN	P49327 homo sapien
14	1018	14.1	4447	1 PKSK_BACSU	P40803 bacillus su
15	997.5	13.8	2511	1 FAS_CHICK	Q12276 gallus gall
16	940	13.0	4427	1 PKSI_BACSU	Q05470 bacillus su
17	882.5	12.2	4273	1 PKSM_BACSU	P40872 bacillus su
18	812.5	8.5	352	1 FAS_ANGAN	P36189 anser anser
19	392.5	5.4	317	1 FAD_BACSU	P71019 bacillus su
20	363	5.0	401	1 NOB4_RHIME	P06631 rhizobium m
21	354.5	4.9	402	1 NODE_RHIME	P06630 rhizobium m
22	344	4.8	308	1 FAD_SALTY	O85140 salmoneila
23	336	4.7	403	1 NODE_RHIV	P04683 rhizobium l
24	332	4.6	308	1 FAD_ECOLI	P25715 escherichia
25	329.5	4.5	467	1 KASA_STRCO	Q02059 streptomyc
26	328	4.5	312	1 FAD_HAEIN	P43712 haemophilus
27	325.5	4.5	402	1 NODE_RHIS3	P07231 rhizobium s
28	324.5	4.5	401	1 NODE_RHIL1	P04684 rhizobium s
29	321.5	4.5	426	1 KASI_STRGA	P16538 streptomyc
30	316.5	4.4	300	1 VAT1_HUMAN	O99535 homo sapien
31	315	4.4	412	1 FAFB_ECOLI	P39335 escherichia
32	308	4.3	414	1 FAFB_VIBHA	P55338 vibrio harv
33	304.5	4.2	293	1 FAD_SYNY3	P73342 synchocyst

34	303.5	4.2	413	1 FAFB_VIBCH	Q94q99 vibrio chol
35	303	4.2	838	1 FAS_MOUSE	P19096 mus musculu
36	302	4.2	420	1 KASI_STRCM	P41175 streptomyc
37	298.5	4.1	327	1 OOR_ECOLI	P28304 escherichia
38	298	4.1	421	1 KASI_STRVN	P16540 streptomyc
39	295.5	4.1	327	1 OOR_SALTY	P40783 salmoneila
40	293	4.1	421	1 FAFB_RHIME	P56902 rhizobium m
41	289.5	4.0	422	1 KASI_STRHA	Q05356 streptomyc
42	285.5	4.0	416	1 FAFB_SYNY3	P73383 synchocyst
43	284.5	3.9	423	1 KASI_STRCO	P23155 streptomyc
44	282	3.9	379	1 VAT1_TORCA	P19333 torpede cal
45	278	3.9	325	1 OOR_PSEAE	P43903 pseudomonas

## ALIGNMENTS

```

RESULT 1
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; 054096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staever M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis."
RL Science 252:675-679 (1991).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevil D.J., Cortes J., Haydock S.F., Leadley P.F.;
RT "6-deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme."
RL Eur. J. Biochem. 204:39-49(1992).
RN (3)
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC
CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHEINES.
CC
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.jeb-sib.ch/announce/> or send an email to [license@jeb-sib.ch](mailto:license@jeb-sib.ch)).

CC EMBL; M63677; AAA26494.1; -;  
DR EMBL; X62569; CAA44448.1; -;  
DR PIR; S23070; S23070.  
DR InterPro; IPR001227; Ac transferase..  
DR InterPro; IPR002085; Adh zn family.  
DR InterPro; IPR000794; Ketocacyl-synt.  
DR InterPro; IPR006163; Pp\_bind.  
DR InterPro; IPR006162; Ppanne\_attach.  
DR Pfam; PF00698; Acyl\_transf; 2.  
DR Pfam; PF00107; ADH zinc N; 1.  
DR Pfam; PF00109; ketocacyl-synt; 2.  
DR Pfam; PF02801; ketocacyl-synt C; 2.  
DR Pfam; PF00550; Pp-binding; 2.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.  
DR PROSITE; PS00606; B KETOACYL SYNTHASE; 2.  
DR TRANSFERASE; Acyltransferase; 2.  
KM Phosphopantetheine; Multifunctional enzyme.  
FT DOMAIN 1 1484  
FT DOMAIN 1485 3567  
FT DOMAIN 27 488  
FT DOMAIN 559 884  
FT DOMAIN 1130 1301  
FT DOMAIN 1397 1467  
FT DOMAIN 1485 1943  
FT DOMAIN 2013 2336  
FT DOMAIN 2383 3066  
FT DOMAIN 3139 3322  
FT DOMAIN 3415 3485  
FT ACT\_SITE 202 202  
FT ACT\_SITE 651 651  
FT BINDING 1430 1430  
FT ACT\_SITE 1661 1661  
FT ACT\_SITE 2105 2105  
FT NP\_BIND 2961 2978  
FT NP\_BIND 3142 3157  
FT BINDING 3448 3448  
FT BINDING 438 438  
FT CONFLICT 480 480  
FT CONFLICT 1241 1241  
FT CONFLICT 2664 2664  
SQ SEQUENCE 3567 AA; 374413 MW; EE284F4738AA0C0 CRC64;

Query Match 29.2%; Score 2107.5; DB 1; Length 3567;  
Best Local Similarity 32.5%; Pred. No. 4.5e-103;  
Matches 610; Conservative 179; Mismatches 530; Indels 559; Gaps 45;

QY 2 ADRIEBAE-DPAIVGACRLPGGVIDLSGFMTLL-EGSRDVGVPWR-WDAAMF 58  
DB 1478 AVYVGAADSEPAIVGICRPPGIGSPBQLWRVLAEGANLVTG-FPADRGWDIGRLY 1536  
QY 59 DDPDAFGKTPVTRASFSDVACDFASFFGISPREALRMDPAHRLLEVCMEALENAIA 118  
DB 1537 HRPDNDNGTGYVDKGFLLTDADDFPGFETPREALAMPQGRKLMETMEANVERAGID 1596  
QY 119 PSALVGTGTGVTGIGPSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGLRPPC 178  
DB 1597 PDALRGDTGVVFVGMNGSQSYQLL--AGEARVVDGYGGLGNSASVLSGRVATYTFGMEGPA 1654  
QY 179 VAVDTAASSSLVAVHLACQLRSGECTALAGVSLMSLSTLVMLSKTRALARDGCKA 238  
DB 1655 LTVDTACSSSLVGLHLMQALRRGCSLALAGVTVMSDPYTFVDFSTQGLASDGRCKA 1714  
QY 239 FSAEADFGGEGCAVVVLKRLSGARADGRRIIIVINGSAINHDGASSGLTPVPGSSOEI 298  
DB 1715 FSAEADFGGEGCAVVVLKRLSGARADGRRIIIVINGSAINHDGASSGLTPVPGSSOEI 298  
QY 299 VLKRALADAGCAASSVGVYEAHGTITLGPDIETIQLNAVYGLGRDVATPLLTGSVKTNL 358  
DB 299 VLKRALADAGCAASSVGVYEAHGTITLGPDIETIQLNAVYGLGRDVATPLLTGSVKTNL 358

DB 1775 VIRQALAAAGVPADVDVVEAHGTETELGDPTEAGALITATGGDRD--RPLRGSVKTNI 1832  
QY 359 GHEPVASGITGLIKVYLSTOHGOIPAHLLAQNLNPRISMGDLRLVTRARTPMDWNTPR 418  
DB 1833 GHTQAAAGAGAVTKVLLAKHGMKLPRLSHADELSPIIDWESGAVEVLRBEVPPACERR 1892  
QY 419 RACVSSFGMSGTNAHVLEEAR-----AATCTPPAPERPAELLVLSARTASALDAQAARL 473  
DB 1893 RACVSSFGMSGTNAHVLEEAR-----AATCTPPAPERPAELLVLSARTASALDAQAARL 473  
QY 474 RDHLETPYPOCLGDVAFSLATTRSAMERHLAAVATSREGRLAALDAAGQTSFGAVRSI 533  
DB 1947 AETHLRPTPELGLDAMWTALATGARFDVAAVLDGDRACVCELDLALAGRPADVAVAP 2006  
QY 534 ADSRGRKLAFLFTGQAGQTLTGMRGLYDVWASAFREAFDLCVRLFNQELRPRLEVMMAER 593  
DB 2007 TSAAPR-KPVLVFPGQAGQWGNARDLBESSVFAEBSMCAELSLPHTWKLLDVRGCG 2065  
QY 594 ASYDAALLDQTAFTQAPLTFEYALALWRSWGBEDELVAHSGISGLVACVAVGSLESD 653  
DB 2066 GPDHHERVD---VLQPVLFISIMVSLAEIWRAGVTPAAVVGSHQGEIAAAHVAGALSLEA 2122  
QY 654 ANFLVARGRLMQALPAGGAMVSTEARPEADVAANAAPHAASVIAAVNAPDOVYVITAGQ 713  
DB 2123 AAKVVALRSQVIRELDDQGMVSVGSRDELETVLRMDGKRVAAVAVNGEPTSVVAGPTA 2182  
QY 714 PVYALIAAAMARGARKTALHVSHPSPMAPPLAEAFGRVAESVSRPSIVLVSNLSGK 773  
DB 2183 ELDEFFPAEAEARMKRRRIAVRYASHPEVARIEDLAEELGITTVAGRSVPLHSTVTGE 2242  
QY 774 ACTDEVSPPGVVWRHAREVRFPADVKAALHAAGAGTFVEGPKSTLL-----GLV 823  
DB 2243 VIDTSAMDASWYRNLRPLVFEQAVRGLVEQCFDFEVEVSPHVLVLMVEETABAHAGAE 2302  
QY 824 PACMPARPALLASSAGADEPATVLEAL--GGLMVGGLSVAGLP--SGGRKRPPLPT 879  
DB 2303 VTCVP-----TLREQSGPHEFLRLRLRAHVHGVG-----ADLRPAVAGGRPAELPT 2349  
QY 880 YPMQRERYW--IDTKADDAARGDRRA----- 903  
DB 2350 YPEHQRFPMPRPAPDVSHLGVKGAEHPLLLAAVDVPGHGAVFTGRLSTDEQPLAEH 2409  
QY 904 -----PG-----AGHD----- 909  
DB 2410 VVGGRTLVGSVLVDLAAAGEDVGLPVLBEVLQRPVLVLAGALLRMSVGAPDESGR 2469  
QY 910 ----- 909  
DB 2470 TIDVHAEDVADLADQMSQAHATGTLAOGVAAPRDEQWPEEDAVRIPLDHIDYGLAQ 2529  
QY 910 -----EV-----EEG-----GAV----- 917  
DB 2530 GYRYGFSFOALRAAMRKDSVYAEVSIADDEBGIAPHVLLDAVQTLISGLALGERGCGK 2589  
QY 918 -----RGDRRSARL----- 930  
DB 2590 LRFAMTVTLHAGATSVRVATPACADAMALRVTPAGHLVATVDSLVRSTGEKWEOP 2649  
QY 931 PPESGRR----- 947  
DB 2650 EPRGEGEGLHALLDMGRVLAEGSTGRVVAADASDLDAVLSRGEPEPPAVLVRYEPEDDPR 2709  
QY 948 ----- 947  
DB 2710 AAARHGVLAALVVRMLEOEBELPGATLVATISGAVTVSDDDSVPERGAAMMGVRCQA 2769  
QY 948 -----FLEID--EKGVL-----DHLVLRV-----TERRAPG 972  
DB 2770 ASEPDRFVLLDTPAEFGMLPAVPDNPQALRGDDVFPRLSPAPSAFTLPACTQRLVPG 2829  
QY 973 -----LGEVEIAYDAAGISFNDVOLALGMVDPDDLPGKRNPELLG 1012  
DB 2830 DGAIDSVAFEPADVEOPURAGEVRVDVVRATGVNFRDVLALLGMYPQKAD-----MG 2881

QY 1013 GCACGIVAVGCVGLVVGOPVIALSAGAFATHVTTSAALVLPFOALSAIEAAMPVA 1072  
 DB 2882 TEAGVAVTAVGVDVDFAFGDRVLGIFQGFAPFIATDHRLLARVDDGSDADAAVPIA 2941  
 QY 1073 YLTAMVADRIRARLOGERVLTHATGVGLAAVQAGVGAHVHATGTGPEKRAYLESU 1132  
 DB 2942 YTTAAVHALDLAGLRAGOSVLTHAAAGVGVMAVALARBAAGAEVLTAG-PAGHGTLRAL 3000  
 QY 1133 GV-RVSDRSRDRFADVAVMTGSGGVVUNLSGELIDSGFNLRSHGHPVLEGRD 1190  
 DB 3001 GUDDEHIASSRRTGFRKFRERTGSGVAVUNLSLTGELDESADLLAEDGVFVENGKTD 3060  
 QY 1191 CVADNOLGURPPLRNLSFSLVLDRCGMLEPPA-----RVRLTEELLGLIAGVF 1240  
 DB 3061 -----LRDGG-----DERG-----RYAPFDGGEAGDRLGELILEVGLGAGSL 3100  
 QY 1241 TPEPIATLPIARVADAFRMAQAQHLKGLVLTG--DPEVOIRIPTHAGAGPSTGRDL 1297  
 DB 3101 DRLPVSAMEIGSAPALQHMRSRGHVGLVLTQPAVDPDGVLI-----TGCTGT 3151  
 QY 1298 LRLIASAPARAALAEFLRQVSVLTPPEIKVGAELFTRLGDSLMAVELRRRIA 1357  
 DB 3152 LGRLL-----LARHLVTEHGVRLH-----LVSRGADAGSGDELRAEIR- 3190  
 QY 1358 SIKLKLTSTFELSTSPNIALLAQNLDPALATL-----SLERVAEN-LR 1400  
 DB 3191 DGGASMEIACPTADBDALSA--LDDGLRPLTGVAHAGVADGLVTSIDEPAVQVLR 3248  
 QY 1401 AGVQNDVSSGADQDWEI 1418  
 DB 3249 AKV-----DAAWNL 3257

## RESULT 2

ERTL\_SACER STANDARD; PRT; 3491 AA.

AC 003131;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-deoxyerythronolide B synthase 1) (DEBS 1).  
 GN ERYA.  
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
 OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetaceae; Pseudonocardiaceae; Saccharopolyspora.  
 OX NCBI\_TaxID=1836;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91220065; PubMed=2024119;  
 RA Donadio S., Steaver M.J., McAlpine J.B., Swanson S.J., Katz L.;  
 RT "Modular organization of genes required for complex polyketide biosynthesis."  
 RL Science 252:675-679 (1991).  
 RN (2)  
 RP SEQUENCE OF 3474-3491 FROM N.A.  
 RX MEDLINE=93231529; PubMed=8386127;  
 RA Donadio S., Steaver M.J.;  
 RT "IS1136, an insertion element in the erythromycin gene cluster of Saccharopolyspora erythraea."  
 RL Gene 126:147-151 (1993).  
 CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-deoxyerythronolide B.  
 CC -1- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN BIOSYNTHESIS.  
 CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORF CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,

CC RESPECTIVELY.  
 CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT), BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR) DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION OF THE FULL-LENGTH CHAIN.  
 CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).  
 CC -1- SIMILARITY: Contains 3 acyl carrier domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL, M63676; AAA26493.2; -.  
 CC EMBL, L07626; AAA26504.1; -.  
 CC PIR, T43231; T43231.  
 CC HSSP, P25715; TMLA.  
 DR InterPro; IPR001227; Acyltransferase.  
 DR InterPro; IPR002198; ADH\_short.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR006163; Pp bind.  
 DR InterPro; IPR006162; Pantane attach.  
 DR Pfam; PF00698; Acyl\_transf\_3.  
 DR Pfam; PF00106; adh\_short; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 2.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 2.  
 DR Pfam; PF00550; pp-binding; 3.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 3.  
 DR PROSITE; PS00705; ACP\_DOMAIN; 3.  
 DR TRANSFERASE; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat; KW Phosphopantetheine; Multifunctional enzyme.  
 DR DOMAIN 1 1972  
 FT DOMAIN 1 1979 3491  
 FT DOMAIN 1 1 375  
 FT DOMAIN 1 414 484  
 FT DOMAIN 1 503 961  
 FT DOMAIN 1 1030 1356  
 FT DOMAIN 1 1611 1794  
 FT DOMAIN 1 1888 1958  
 FT DOMAIN 1 1979 2441  
 FT DOMAIN 1 2507 2854  
 FT DOMAIN 1 3055 3237  
 FT DOMAIN 1 3334 3404  
 FT ACT SITE 145 145  
 FT ACT SITE 145 145  
 FT BINDING 447 447  
 FT ACT SITE 677 677  
 FT ACT SITE 1128 1128  
 FT NP\_BIND 1614 1660  
 FT BINDING 1921 1921  
 FT ACT SITE 2148 2148  
 FT ACT SITE 2598 2598  
 FT NP\_BIND 3058 3104  
 FT BINDING 3367 3367  
 SQ SEQUENCE 3491 AA; 365022 MW; 682BFC3C90FA8C4 CRC64;

Query Match 26.7%; Score 1925.5; DB 1; Length 3491;  
 Best Local Similarity 35.6%; Pred. No. 1.7e-93;  
 Matches 347; Conservative 184; Mismatches 601; Indels 203; Gaps 39;

QY 4 RPIERAEDPIAIVAGSCLPGVLDLSGFWTLGSRDVTGVRAER-MDAAMPDPDP 62  
 DB 497 RNEAPGEPVAVVMAACRLPGVSTPEEFELLSEGRDAVAGLPLDRQMDDSL.FHPDP 556  
 QY 63 DAPGKTPVTRASFSDVACFDSFGISPRELRNDPAHRLLEVCWELEMAALAPSL 122  
 DB 557 TRSGTAHORGGFLTEATAFDPAFFGMSPRELAVDPOORLMLLSWELELAGIPPTSL 616

```

QY 123 VGTETGFIGIGPSEYEALPQATASAEIDAHGGLAGTAMPVSGAGRIYALGKPCVAVD 182
D 617 QASPTGTFVLLIPOEYEPRL--AEGGEGVGYLMTGTTTSVAGRIAYITGLGSPALSYD 674
QY 183 TAYSSILVAHVLAQCSLRSECTALAGVSLMSPSTLYWLSKTRALADGRCARSAE 242
D 675 TACSSSIVAHVLAQCSLRSESSILAMAGVLTVPMTPLMVDPSMNSLADGRCARSAE 724
QY 243 ADEFGEGGCAVVVRLKSGARADGRIILAVIRGSAIINHDSAGSLVPPGSGOEITLKR 302
D 735 ANFGNAEGAGMLLERLSDARRNGHPVLAVLKTAVNSDGSAGSLAPAGKQVRYTQQ 794
QY 303 ALADACGAASSVGVYEAHGTGTTLDPIETIOALNAVGLGRDVAIPLIGSVKTNLGHE 362
D 795 ALAESGLGRADIDAVEHGTGRUGDPIEARALFEAY--GRDRQPHLOSIVSINLHGTQ 852
QY 363 YASGITGLKLVLSLOHQOI PAHLAQAOLNPRISWGDRLTVTRARTPMDMTPRRAGV 422
D 853 AAGVAGVIMVILAMRAGTLPRTLHASERSKEIDMSGAIISLDEPEPMPAGARPRRAGV 912
QY 423 SSTGMSGTNAHVLEAP-----AATCTPPAPRRPAELVLSARTASALDAQARL 473
D 913 SSFGISGTNAHAIIEEAPQVEGEGRYEAGDVAP-----WLSASASBGLRAQARL 964
QY 474 RDHLETYPSQCLGDAFSLATTSAMEHRLAVALTSREGLRAALDAQAQOTSFGAVRSI 533
D 965 AALHREHPGDDPRDIASLATGSAALPHRAAFAPVDSALRLVLDIGLATNAGCAVGT- 1023
QY 534 ADSSRGKLAFLFPGQAGQTLGMRGLYDVWSAFREAFDLCVRLFNOELDRPLREVMKAP 593
D 1024 -SRAOQRAVFPFGQGMAGMAVDLIDTSPVFPAALRECADALDEPHLDEVAIFPLAAEA 1082
QY 594 A--SVDAAL-LDQTAFTOPALFTEFVALALNRSWGPBEPVLVNGHSIGELVAACVAGVS 650
D 1083 ARBEQDAAUSTERTDVQVPMFAVMSLASMMRHGVEPAVAVIGHSGEELAAACVAGALS 1142
QY 651 LEDAVFLVAARGRLMOALPAGAMVSIIEADVAALVAAPHAASVSIAAVNAPDOVIAG 710
D 1143 LDDAARVVALRSVIAITMPGNKGAISIAAPAGEVRAIIGR---VELAIVNGPRSVVAG 1199
QY 711 AGCPVHAIAAAMAAAGARTALHVSASFPLAEMLEAFGRVAESVSYPSPSTIV-LVSN 769
D 1200 DSDLDLVLVASCCTECIRAKRLAVDVYSHSHVETIIDAHAELGDEPHLPGFVPFST 1259
QY 770 LSKACACDEVSSPGYWRHAREVVRFPADGVKALHAAGAFVEVEXPRSTLGLVPACMPD 829
D 1260 VTGWTOPDELDAGYWRNLRRTVRFADAVRALBQGYRTFLEVSANPILTAALIEETGDC 1319
QY 830 ARPALLA--SSRAGDEPATVLEALGGLMAVGLVSWAGL-FPSGGRVPLPTYMQRER 886
D 1320 SGADLSAISHLRGDSGLADFGELISRAFAAGVAVDMESVHLTGARVRPLPTYPQORER 1379
QY 887 YWIDTKADDAARGDRAAPGAGHDEVEEGAVR-----GGDRRSARDL----- 928
D 1380 VMLEPK--PAARRS-----TEVDVSALRYRIEMRPICGAGEPARLDTGWLVAKYAGT 1429
QY 929 HPPESGRKKVBAAGRPRLIEIDEPGVLHLVLTERRABELGVE-----IADAA 983
D 1430 ADETSTAREALBSAGKARVELVVDARCGDELA-----ELKASVGEVAGVLSILAIDEA 1484
QY 984 GLSFNDVALLGMVDDLP-----GKPNPPLLIGEGCAGRIYAVEGVNGLV--V 1031
D 1485 --EPEEARPLALASLADTLISLVQAMVSAELGCLMTVESAVAGPFRFVANAAGALMGV 1542
QY 1032 GQPVYIALSAGAFATHVTTSAAVLVPRQALSAIE-----AAAMPVAYLTAWVALDR 1082
D 1543 GRVIALENPAVWGGLVVPAGSVAEELARHLLAAVAVSGAGEDQALRADGVYGRWRVAAA 1602
QY 1083 IA---RIQGERVLIHAATGVGLAAVQMOHVGAEEVHA-----TAGTEPKRYALE 1130
D 1603 PATDDEKPKPTGTVLVTGTGTGGVGOQIARWLARRAP--HLIVSRSGPDADAGLVLELE 1661
QY 1131 SLGVRY-----VDSRSRDRFVADVVRAMTGEVGVVL-----NSLSGEL 1169

```

```

D 1662 ALGARTTVAACDVTDRES-----VRELLGGICDDVPLAVFHAATLDDGTVDLTGER 1715
QY 1170 IDKSFYLLASHGRFVELGRDVC-ADNOUGLAPFLNLSFSL-----VDR 1214
D 1716 IERA-----SAKVLGARNILHETRELDLTAFLVFSFPASAFGAPGLGTAAPGNAVLD 1768
QY 1215 GMMLEL-----PARVA-----LLEELGLIA-----AGVFTPPP----- 1244
D 1769 GLAQGRSGLPATANAKMTWAGSSWAE---GAVADRFRRHGVIEMRPETACALQNALD 1825
QY 1245 ---IATLPIARVADAFRSNAQAQHLKVLVTLGDEPVQIRIPTHAGAGSTGDRLLDLR 1301
D 1826 RAEVCFIVIDVRWDRFLAYTAQRPTRLDEIDAR---RAAPQAPAEPRVG-----AL 1876
QY 1302 ASAAPARAALAEAFRTQVSQVL-RTPPEIKYGAELFTRLGMDSLMAVELNRIRIASLK 1360
D 1877 ASDPAREBEALFELVRSNAALVGHASAEERPADAFAELGVDLSIALELRRLGAATG 1936
QY 1361 LKSTFELSTSPNIALAQNLDALATALSLEVA 1395
D 1937 VRLPTTVDPHPDVRTLAHLAAELGATGAEGQA 1971

RESULT 3
OL56 STRAT
ID OL56 STRAT STANDARD; PRI: 3519 AA.
AC Q07017;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Oleandomycin polyketide synthase, modules 5 and 6.
GN ORF.
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150470; PubMed=8107683;
RA Swan D.G., Rodriguez A.M., Vilches C., Salas J.A.;
RT "Characterisation of a Streptomyces antibioticus gene encoding a type
RT I polyketide synthase which has an unusual coding sequence.";
RL Mol. Gen. Genet. 242:358-362(1994).
CC - FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC LACTONE RING.
CC - COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
CC - SIMILARITY: Contains 2 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: 109654; AAA19695.1;
CC PIR: S43048; S43048.
CC HSSP: P25715; IMLA.
DR InterPro: IPR001227; Ac transferase.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR006163; Pp bind.
DR InterPro: IPR006163; Ppantne_attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl transferase.
DR Pfam: PF00109; ketoacyl-synt. 2.
DR Pfam: PF02801; ketoacyl-synt. C; 2.
DR Pfam: PF00550; pp-binding. 2.
DR Pfam: PF00975; Thioesterase; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 2.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 2.
DR PROSITE: PS50075; ACP_DOMAIN; 2.

```



KM Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;  
KW Phosphopantetheine; Multifunctional enzyme; Repeat.  
FT DOMAIN 1 3519  
FT DOMAIN 2 501  
FT DOMAIN 3 569  
FT DOMAIN 4 880  
FT DOMAIN 5 1200  
FT DOMAIN 6 1487  
FT DOMAIN 7 1561  
FT DOMAIN 8 1686  
FT DOMAIN 9 2220  
FT DOMAIN 10 2541  
FT DOMAIN 11 3038  
FT DOMAIN 12 3215  
FT DOMAIN 13 3270  
FT DOMAIN 14 3519  
FT ACT SITE 210  
FT ACT SITE 660  
FT NP BIND 1203  
FT BINDING 1524  
FT ACT SITE 1859  
FT ACT SITE 2311  
FT NP BIND 2859  
FT BINDING 3178  
SQ SEQUENCE 3519 AA; 368561 MM; 41AE78AAAE61F86 CRC64;  
Query Match 26.2%; Score 1889; DB 1; Length 3519;  
Best Local Similarity 34.0%; Pred. No. 1.4e-91;  
Matches 541; Conservative 191; Mismatches 595; Indels 264; Gaps 44;  
9 AAEDIIATVGCACRLPGGVIDLSGFMTLESGRDTYGRVPAER-DMAAAMPDDPARCK 67  
Db ASREAIATATMSCRGGGIDSPEDLWFLAEGRDAAVAGLPEDRGMDLALYHPDENPCT 1743  
Qy TPVTRASFSLDVACDFASFFGISPREALMDPAHRLLEVCWEALENNAIAPALVGET 127  
Db TVVRGAFRYDAAGDAGFFGISPREALMDPQRLLETSMELFERADIDYTVRGTA 1803  
Qy 128 GVFTIGTIGSEY--EALDPATASABIDAAGLGTWPSVAGARISYALGLRPPCAVDTA 184  
Db 1804 GFIFAGHGYGPPDKRAPEESVAGYLL-----TGTAVALSGRISTYFGLGPAVTVD 1858  
Qy 185 YSSSLVAHVLAQCSLRSGECSTALAGVSLMSPSTLVWLSKTRALARPGRCKASAE 244  
Db 1859 CSSSLVAHVLAQALRGECSTALAGVAVMSTPDAVEFSSROQGARGRCKAPAAAD 1918  
Qy 245 GFGREGCAVVVVKELSGARADGDRILLAYIRGSAINHDGSSGLTVPNSSOEIVLKRAL 304  
Db 1919 GKGWEGSVSLLEHLSDRRIGHVLAIVRGSAVNOGASNGLAHPNGSQORIVIRAL 1978  
Qy 305 ADAGCAASSVGVVEAHGTGTTLGDPRIEQALNAVYGLGRDVAATPLLIGSVKTNLGHPEYA 364  
Db 1979 ADAGLAPADVDAVEAHGTGTRLDGPIEAQALATVGGGRAGRPVWLGSVKSNIGHTQA 2038  
Qy 365 SGITGLLVVLSLQHGQIPAHILHAQALNPRISMGDLRLTVTAARIPW-PDMTPPRAGVS 423  
Db 2039 AGVAACVMKQVNLGRGVVPKTLHVDSPSHVMSAGAVELLTEERPEWEAEARLRAGIS 2098  
Qy 424 SFGMSGTNAHVLEEAAPATCTPPAPERPAELL-----VLSARTASALDAQAARLR 474  
Db 2099 AFGVGTNAHVLEEAAP-----EPEPEPCTRYVAAGDLVVPVVGSRDVGALREDAARLA 2154  
Qy 475 DHLETFYPSQCGDVAFSLATTRSAMEHRLAIVAASTRREGRLAALDAAAGQTSPPGAVRSIA 534  
Db 2155 AHVSTGTAGVV-DVGMSTLVAATRSVEHRAVMVGTDLDSWAGSLAGPAAGGVVPGVAGVA 2213  
Qy 535 DSSRGKLAFLFTGGCAQTLGKMGRLYDVMSAFREAFDLCVRFNDELDRPLREVMABRA 594  
Db 2214 PAEGRRVVFPPGQSQVGMMAAGLLDACPVFAEVAECBAALVDEVGMSLVEVLQGR-- 2271  
Qy 595 SVDAALLDOTAFTOPALFFFEYALALMRSWGVPEBELVAGHSIGELVAACVAGVFSLEDA 654  
Db 2272 --DAVVLGRVDVQALMAVMVSLARTMYRYYVEVRAAVYGHSGEIIAAACVAGGSLADG 2329  
Qy 655 VFLVARGRLKQALPAGGAMVSIIEAPEADVAAVAAPHAASVIAAVNAPDOVVIAAGAP 714

Db 2330 ARVVLRSRARIALIAGGGMVSLSLPAGRVRITMLDTYGGVRSVAAVNGPSSVTVSGDYQA 2389  
Qy 715 VAIAAAMAARGARTKALVSHAFSPPLMAPLEAFGRVAASVSRRPSIVLVSLSGCA 774  
Db 2390 LDELIIAGGEREVRKARRVPDIASHSAQMDQIRDLLEALADITQDSSVPPFFSTVTADW 2449  
Qy 775 CTDEVSSPGYVWRHAREVVRFPADGVKALHAAGAGTFVEVGPSTLLGLVPACMP--DARF 832  
Db 2450 LDTTALDAGYWFNTNREIVRFGAEVAGLVAGQMGAFVSCSPHPVLVPGIEQLDALDQNA 2509  
Qy 833 ALLASSRARGDEPA--TVLEALGILMAVGLVSYNA-----GLFPGSGRRVPLFTYMQORR 886  
Db 2510 AVLGSLR--REGGIDRLTLTSLAEAFVQGVDPDWTHTAEVTP--RTVDLPTYPFORQR 2564  
Qy 887 YVIDRKADDAARG--DRRAPAGHDVEEGAVR-----GGDRSARLDHPPEESRREV 940  
Db 2565 FMLDSSPASSANGVGEADAMITWDVEREDSVAAVEELGIDDEALHTVLPALSSRRRRV 2624  
Qy 941 E-----AAGDRPFLLEIDE-----PGVLIDLVLVTERRAPGLGE 975  
Db 2625 EHRRLODMRYVEMKFPFAALDEVLGGGMLFVPPRGGLADDGVAVVAVAVTARG--GE 2680  
Qy 976 VEI-----AVDAAGLSFNDVOLALGMP--DDLPGKPNPPLLIGGECA 1016  
Db 2681 VSVVELDPTRPDRRAYAEAVAGRGVS-----GVVSFLSWDDRRHSEH-FVVPAGILAA 2731  
Qy 1017 GRIVA-----VGEQVN-GLVVGQPVYA--LSAGAFATVTTTA-----ALVLPKPO- 1059  
Db 2732 SLVLAQALVDLGRVGEGRMLVTRDAVAVAGPSDAVADPQAOVGFGRVLGHEHEL 2791  
Qy 1060 -----ALSAIEAAA-----MPVAVLYATWYALDRIRAL----- 1086  
Db 2792 MGLGLDLVEAPPEPGSTCDHTYADLATVVASAGEEDQVAARGSGVW--VRLVBAVAVDG 2849  
Qy 1087 -----QGEERYLVIAATGCGVGLAAYQAOHVGAEVH-----ATAGTEPKRAYLESL 1132  
Db 2850 GGGGWRPRTVLYVTGGLGAHTARMLVGGAD-HVVLVSRGGSAAGACDLVRELEGL 2908  
Qy 1133 GVRVYSDRSRPFVD--VRAWTG--GEGVDVYVNSLS----- 1166  
Db 2909 GGARVSVRACD--VADRVALLSLSDGEPYTAFFHAAGVPOSTPLAEISVGEADAVMAA 2966  
Qy 1167 -----GELID-----KSFNLSRSHGRFVELGKRDCVA----- 1193  
Db 2967 KVAAGVNLGELVDPGGLAFLFSSNAGWMSGGCAVYAAANAFILDALAVRRGVLGART 3026  
Qy 1194 -----DNQGLRPFLRNLSFSLVDLRGMLEERPVRVALLLELLGLIAAGVFTPPP 1244  
Db 3027 SVAMGMWAGEGMASVGAARELS-----RRGVRAWDPERAVAVMAVADVGRGAFV--A 3077  
Qy 1245 IATLPIARVADAFFSMQAQHGLKVLVTLGDPVEQIRI-----PTHAGAGPSTGDRULL 1298  
Db 3078 VADVMEFVTGFAFSARPRPLISDL-----PEVKTALRNOBOEQHAPV-PEDRSQQL 3130  
Qy 1299 DRLASAPARAARAALEAFITQVSOVLRTPEIK-VGAELPTRLGMDSLMAVELNRIEA 1357  
Db 3131 RRLMSLSPAGGEAEVLYKVRTEAAVAVLGHSAQDVPAARAKELGFSBLTAVQLRNRLAA 3190  
Qy 1358 SLKULSTFLSTSPNIALLAQNLIDALATA 1388  
Db 3191 ATGTRLPASAVFDHPHAALARMLLAGMRHA 3221  
RESULT 4  
ERY3\_SACER STANDARD; PRT; 3172 AA.  
AC Q03133; Q54097; Q99270;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.3.1.94) (ONF 3) (6-deoxyerythronolide B synthase III) (DEBS 3).  
GN ERYA.

```

OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=91043075; PubMed=2234082;
RA Cortes J., Haydock S.F., Roberts G.A., Bevilitt D.J., Leadlay P.F.;
RT "An unusually large multifunctional polypeptide in the erythromycin-
RL Nature 348:176-178(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Steaver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RL biosynthesis."
RL Science 252:675-679 (1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevilitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme."
RL Eur. J. Biochem. 204:39-49(1992).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- PATHWAY: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- BIOSYNTHESIS.
CC -1- BIOSYNTHESIS. IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -1- BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETO REDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: CONTAINS 2 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X56107, CA39583.1; -
DR EMBL, M63677, AAA26495.1; -
DR EMBL, X62569, CAA4449.1; -
DR PIR, S13595, S13595.
DR PIR, S22012, S22012.
DR HSSP, P00101, ICCH.
DR InterPro, IPR001227, Ac_transferase.
DR InterPro, IPR002198, ADH_short.
DR InterPro, IPR000794, ketoacyl-synt.
DR InterPro, IPR006163, Pp_bind.
DR InterPro, IPR006162, Ppantn_attach.
DR InterPro, IPR001031, thioesterase.
DR Pfam, PF00698, Acyl_transf. 2.
DR Pfam, PF00106, adh_short. 1.
DR Pfam, PF00109, ketoacyl-synt. 2.
DR Pfam, PF02801, ketoacyl-synt_C. 2.
DR Pfam, PF00550, pp-binding. 2.
DR Pfam, PF00975, thioesterase. 1.
DR PROSITE, PS00012, PHOSPHOPANTETHEINE. 2.
DR PROSITE, PS00606, B_KETOACYL_SYNTHASE. 2.
DR PROSITE, PS50075, ACP DOMAIN. 2.
DR TRANSFERASE; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KM Phosphopantetheine; Multifunctional enzyme.
FT DOMAIN 1 1485 3172
FT DOMAIN 2 37 484
FT DOMAIN 3 554 878
FT DOMAIN 4 1116 1298
FT DOMAIN 5 1394 1464
FT DOMAIN 6 1488 1954
FT DOMAIN 7 2021 2335
FT DOMAIN 8 2555 2735
FT DOMAIN 9 2821 2891
FT DOMAIN 10 2926 3172
FT ACT SITE 643 643
FT ACT SITE 199 199
FT NP_BIND 1118 1164
FT BINDING 1427 1427
FT ACT SITE 1661 1661
FT ACT SITE 2112 2112
FT NP_BIND 2557 2605
FT BINDING 2854 2854
FT CONFLICT 231 231
FT CONFLICT 240 240
FT CONFLICT 289 289
FT CONFLICT 493 493
FT CONFLICT 493 517
FT CONFLICT 510 510
FT CONFLICT 513 513
FT CONFLICT 525 525
FT CONFLICT 536 536
FT CONFLICT 547 551
FT CONFLICT 553 553
FT CONFLICT 673 673
FT CONFLICT 716 716
FT CONFLICT 734 736
FT CONFLICT 896 896
FT CONFLICT 898 994
FT CONFLICT 1108 1116
FT CONFLICT 1124 1126
FT CONFLICT 1132 1132
FT CONFLICT 1192 1192
FT CONFLICT 1194 1194
FT CONFLICT 1277 1278
FT CONFLICT 1385 1390
FT CONFLICT 1485 1485
FT CONFLICT 1518 1518
FT CONFLICT 1601 1601
FT CONFLICT 1724 1725
FT CONFLICT 1732 1732
FT CONFLICT 1733 1743
FT CONFLICT 1762 1762
FT CONFLICT 2252 2252
FT CONFLICT 2275 2277
FT CONFLICT 2408 2408
FT CONFLICT 2420 2421
FT CONFLICT 2443 2444
FT CONFLICT 2596 2596
FT CONFLICT 2609 2609
FT CONFLICT 2715 2722
FT CONFLICT 2754 2754
SQ SEQUENCE 3172 AA; 331474 MW; DBD5094877DDDF CRC64;
Query Match 26.0%; Score 1875; DB 1; Length 3172;
Best Local Similarity 34.6%; Pred. No. 6; 9e-91;
Matches 549; Conservative 184; Mismatches 602; Indels 254; Gaps 48;

```

QY 4 RPIERAEDPPIAVGASCRLLPGCVIDLSCFWTLLEBSSRDVTGVRPAERWDAAMPDPDD 63  
 1 REVERHAGEPIAVIAGMACRFPDGDVDSPESEFVEVSCGDALIAEAPDR-----GM-EPDSD 85  
 QY 64 ABGKTPVTRASFUSDVACDASFFGISPREALRMDPAHLLLEVCHEALENNALIASALV 123  
 Db 86 A-----RLGGMILAAAGDADGAFGGISPREALAMPQQRIMLEISWEALERGHDPVSIR 139  
 QY 124 GTEGVFIGIGPSEYEALPOATASAEIDAHGGLGTMPSVGAGRISYALGSPCAVDT 183  
 Db 140 GSATGVFTGVGTVDVGPDPDE--APDEVGIVGTGTASSVASSGRVAVCGLGSPMTVDT 197  
 QY 184 AVSSSLVAVHLACOSIRSECESTALAGVSLMSPETLVMLKTRALARDGRCKAFSAA 243  
 Db 198 ACSSGLTALHAMESIRDECGALAGVTVMSPPCAFTFERSQGLAADGRCKPSSKA 257  
 QY 244 DDFGEGGCAVVLKRLSGARADGRILAVIRGSALNHGASGLTVPNGSQEYLKKA 303  
 Db 258 DDFGLAEGAGVVLORLSAAREBGRVLAIVLAGSANOQASNGILAPSGPAQORVIRRA 317  
 QY 304 IADAGCAASVGVVEAHGTTGLGDPRIEIOALNAVYGLRDVATPLLIGSVKTNIGHPEY 363  
 Db 318 LENAGRADVDVYEAHGTIRLDPIEVALHLLSTYGAERDPDDPLMISVKSNIQHTOA 377  
 QY 364 ASGITGLKVLVLSLOHQIIPAHLHAQALNPRIISMGDLRLTVTRARTPMPDMNTPRAGVS 423  
 Db 378 AAGVAGVMAKVALALRRGEMPTLHDEPSPQIEMDGNVSVSQAASWAGERRPRAGVS 437  
 QY 424 SFGMSGTNAHVLEAPAACTTPRAPERPAELIVLSARTASALDAQARLRDHLLETYSQ 483  
 Db 438 SFGISGTNAHVLEAPADEPAPDSGPVPLVLSGRDEQAMRAQAGRLADHLAPEPAN 497  
 QY 484 CLGDVAFSLATRSAMEHLVAATSR-----GIRALDLAAAGQTSRPAVRSIAOSSRK 540  
 Db 498 SLRDGTFTLATASAEHRAVVGDRDELGLGRAVADRRIADRYATGQ-----GPPSPRR 553  
 QY 541 IAFLEFGGCAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLEVM--MAEPASVDA 599  
 Db 554 VAMVFPGGQAOQGMARBDLRESQVADSIKDCERLAHVHWSLTDLSGARP----- 607  
 QY 600 LLDQTAFTOPALFTEYALALALMRSGVEPELVAGHSIGELVAACVAFSLIEDAVFLVA 659  
 Db 608 -LDRVDVVOPALFAVWVSLAALMRSHGVPAVVGSHOGEIAAAHVAAGALTLEDAAKVA 666  
 QY 660 AGRILMOALPAGAWVSIAPADVAANAAPHAASTIAVANP--DOVVIAAGQGVNHI 718  
 Db 667 VASRVLRIRGGQGNASFGELGTEQOAEIRIGRFAGALSTIASVNGPSVVVAVAGEGSLDEL 726  
 QY 719 AAMAMARGARTALHVSFAHSPILMAPMLEAFGRVASEVSYSRPSIVLSNLGKACTDE 778  
 Db 727 IAECEEAHKARRIPVDYASHSPQVESLREBELTELAGISPVASADVALSTTTGQIDTA 786  
 QY 779 VSSPGVWHAAREVVRADGVKALHHAAGTVEVGPCKSTLLGLVPACMPDARA----- 833  
 Db 787 TMDTVWYVYANLREQVYFQDARTQLAEAGFDFAVEVPHVLTGVIATLDSALPADAGAC 846  
 QY 834 LLASRAGDEPATVIEALGILMAVGVLSMAGLFFSGGRRPVLPPTYMQRERYWDTQA 893  
 Db 847 VVGTLRRDGGGLADPHTALGEAVAGVEVDWSPAF--ADARPVELPYPFO--RYMPLIPT 903  
 QY 894 DDAAR---GDRAPGAGHDEVEGCAVGRDRRSARLDHP--PPE--SGRREVEAAGDR 946  
 Db 904 GGRADDEDDWRYQVVMREAEWESASLAG---RVLLVTGPGVPSLSDAIRSGLEBSGAT 960  
 QY 947 PRLLEID-----EPGVLDHL---VLRVTERRA--PGLGEVEI--AVDAAG----- 984  
 Db 961 VLTCDVESRTIGTALAEADDTALSTVGAAPVHGAVDPSIDALALVQALGAAGVADL 1020  
 QY 985 --LSEFNDVOLAIQMVDDDLPGKENPULLIGGECAGIVVGE-----GVNGIIVQGPVIA 1037  
 Db 1021 WVLTRAAVVADELVD--PAQ---AMVGG--LGVVIGIEQGRNGGLVDL----- 1065  
 QY 1038 LSGAFATHTVTSALVL--PRPOLSAIEAAMPVAVLTAWYALDRIBLQIP--GERVLH 1095

Db 1066 ---DADAASIRSLAVALADPRGEEQVIRADGICKVARIIVPAPARARTHPLEBLACTVLT 1122  
 QY 1096 AATGCVGLAAVOMQAHVGAENHA-----TACTPERKATLEBLGVRY-----VSDS 1140  
 Db 1123 GGTGIGIHLARLWMLRSSAE--HLVLLGRGADAPASBELREBLTALGVTTIACDVADR 1181  
 QY 1141 RSDRFVADVRAVMTGEGVDVVLN-----SLSGELDKSFNLLRSRQRFVELGRDCVADNQ 1196  
 Db 1182 ARLEAVLAAEAABERTVSAMHAGVSTSTPLD-----LTAEPFEI--AD-- 1227  
 QY 1197 LGLRPLENLSFSLVDLRGMMLERPARVALLIEBLGLIA---AGVFTPEPIATLPAR 1252  
 Db 1228 -----VKVRGTV-----NDELCPDDDAFVLFSSNAGVSGPGLASVAAAN 1268  
 QY 1253 V-ADAFRMAQA-----QHLGKVLVLTGP-----EVQIRI 1282  
 Db 1269 AFLDFAPARAREGAPVTSIAGWLAGQNMAGDEGEYLRSGCLAMPDRAVEBLHTYL 1328  
 QY 1283 PTHAGAGPSTGDRD-----LIDRLASA-----APARAAA----- 1312  
 Db 1329 -DHGOTSVSVYDMDRRFEVELTARARHPLPDEIAGATAEARSSEGRPALAQLAALCD 1387  
 QY 1333 -----LEAPLRQVSOVL--RTPEIKVGAELFTLGMDSLMAVELRNRIEASLKLUST 1365  
 Db 1388 GREHEHLHLIRAEVAAVILGHGDDAIDRDRAFRDLGFSMTAVDLRRRLAVALTVGREAA 1447  
 QY 1366 TFLSTSPNIALAQLLALATLSLERYAAE-----NLBAQVON 1405  
 Db 1448 TVFPHPTITRLADHYLERLVGAABAEQAFLVREVPKADDPPIAIVGMACRPPGVAN 1507  
 QY 1406 -----DFVSSG-----ADDMEIIL 1421  
 Db 1508 PGLMEFTVGGDAVTEPTGRMDLAL 1536

RESULT 5  
 MCSA MYCBO  
 ID MCSA MYCBO STANDARD; PRT: 2110 AA.  
 AC 002251;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Myceroic acid synthase.  
 GN MAS.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxId=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=92406887; PubMed=1527058;  
 RA Mathur M., Kolatukudy P.E.;  
 RT "Molecular cloning and sequencing of the gene for myceroic acid  
 RT synthase, a novel fatty acid elongating multifunctional enzyme, from  
 RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guérin.",  
 RL J. Biol. Chem. 267:19388-19395(1992)  
 CC -I- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH  
 CC METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO  
 CC FORM MYCOCEROSYL LIPIDS.  
 CC -I- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE.  
 CC -I- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL  
 CC ARRANGEMENT.  
 CC -I- SUBCELLULAR LOCATION: Membrane-associated.  
 CC -I- SIMILARITY: Contains 1 acyl carrier domain.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
EMBL: M95808: AAA25369.1; -  
DR PIR: B44110: B44110.  
DR HSSP: P73283: 1ESM.  
DR InterPro: IPR001227; Ac transferase.  
DR InterPro: IPR002085; Adh\_zn\_family.  
DR InterPro: IPR000794; ketoacyl-synt.  
DR InterPro: IPR006163; Pp bind.  
DR InterPro: IPR006162; Phantne\_attach.  
DR Pfam: PF00698; Acyl\_transfer\_1.  
DR Pfam: PF00107; Adh\_zinc\_N\_1.  
DR Pfam: PF00109; ketoacyl-synt\_1.  
DR Pfam: PF02801; ketoacyl-synt\_C\_1.  
DR Pfam: PF00550; pp-binding\_1.  
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.  
DR PROSITE: PS00066; B\_KETOACYL\_SYNTHASE; 1.  
DR PROSITE: PS50075; ACP\_DOMAIN; 1.  
KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;  
Transferrase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.  
FT DOMAIN 1 430  
FT DOMAIN 533 852  
FT DOMAIN ? ?  
FT DOMAIN ? ?  
FT ACT\_SITE 2026 2096  
FT ACT\_SITE 177 177  
FT ACT\_SITE 623 623  
FT NP\_BIND 1561 1578  
FT NP\_BIND 1765 1780  
FT BINDING 2059 2059  
SQ SEQUENCE 2110 AA: 225577 MW: C658215D7155300 CRC64;

Query Match 25.7%; Score 1852.5; DB 1; Length 2110;  
Best Local Similarity 29.5%; Pred. No. 6.Se-90;  
Matches 524; Conservative 176; Mismatches 519; Indels 555; Gaps 30;

QY 13 PIAIVGASCRLLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAMFPDDPAKGTPTVTR 72  
DB 7 PVAVIGMGCRLPGGINSPPDKLMESILRGDDLVTETIPDDRWDADDYDPEGVGRSVSRW 66  
QY 73 ASFLSDVACPDASFPGISPREALRMDPAHRLLEVCHEALENMAIAASALVGTETGFIG 132  
DB 67 GGFLLDVAGFPAEFPGISSEBATSIDPOQLLETSEALEHAGLDASLAGSSTAFTG 126  
QY 133 IGPSEVEAALPQATASAEIDAHGGL-----GTMPEVGAGRTSYALGLRGPCCAVDTAY 185  
DB 127 LTHEDYIVLTFTA-----GGLASPYVTGTGNNVVASGRITAHITGLGHPATPTTAC 177  
QY 186 SSSLVAVHLACQSLRSGECSTALAGVSLMSPETLWLSKTRALARDGCKAFSAEADG 245  
DB 178 SSGIMAVHLACRSIHDERDLALAGCAVLLPFAVCVAASQGMWLSSTGRCHSFDADAG 237  
QY 246 FGRGEGCAVVVLRKLSGARADGRIILVIRGSAIINHGAASSGLVTPMGSGOEIYLKALA 305  
DB 238 FVNSEGCAWMLKRLPALRDGNRIFAVVGSTATNQGRTETTLMPSEDAQVAAYRALA 297  
QY 306 DAGCAASVGVVEAHGTGTTGLDPIETIOALNAVGLGRDVATPLLISVKNLGHPEYAS 365  
DB 298 AAGVQPTVGVEHGTGTGPIGPIEYRSLARYGAG-----TPGALSSAKSMGHSSTASA 353  
QY 366 GITGLKLVLSLQHGQIPAHLLAQALNPRISWGDRLITVTRATTPWDMN--TPRRAVGS 423  
DB 354 GTVGLIATILSLRIGVVPRLHFNRLLPDELSDVETGLFVPOAVTPMPNGDHTPKRAVVS 413  
QY 424 SFGSSGNNAVHVEEAATCTPRAPRPAE-----LLVLSARTSALDAQAARLDHLET 479  
DB 414 SFGMSGNNAVHVEEAAPASAPSSPGDAEVGRPLFMELSTSSDALARQTRALQATVVEE 473  
QY 480 YPSQCL--GDVAFSLATTRSMERHLAVAASTR-----EGLRAALDAAAGQTSFGAVRSI 533  
DB 474 H-QDCVAASDLAYTLARGRRRPVYRTAVVAAANLPELYEGRLREVADGALYDAAG----- 527  
QY 534 ADSRGKLAFLFTGQAGQTLGMGGLYDVVSAFREAFDLCVRLFNQELDRPLREVMVAEP 593

DB 528 -HGDRGPV-WVPSGQGSQWAMGTOLL-----ASEPVFAATI-----AKLEPVIAAS 573  
QY 594 A-SVDAALLDQFAFT-----QPALTFPEYALAA-LWRSGVPELVAHSIGELVACVA 646  
DB 574 GRSVTAITAAQQTIVTVIDKQVAVFAVQVALAATMTQTYGVRRGAVGHSMSGSAVAVA 633  
QY 647 GVPSLEDAVFLVAARGRLMOALPAGAVMSIEAPEADV-AAAVAPHAASVSIANAAPQ 705  
DB 634 GALTLEDAARVICRFSKMTMTRIAGAGMGSVLEPAQVNSELMARIGIDVVVSVASPOS 693  
QY 706 VVIAGAGQVYHAIAMAAARGARTKLYSHAFHSTLAMPMLAEFRVAVESYRRPSIV 765  
DB 694 TVVIGGSTIVTRDLIAEQRQDVAVAREVAVDVASHSPVDPIDDLALALADIAPMTPKVP 753  
QY 766 LVSNLSGKACTDEVSFGYVVRHAREVRFADCVKLLHAAGAGTEVEVGPSTLLGLVPA 825  
DB 754 YVSATTFDREBQVCCGATVVDNLRKTTQVPAALVQALMEDGTRVFADVSPHLLTHAVQ 813  
QY 826 C-MPDARPALAASSRAGDEPATVLEALGLWAVAGLVSMAGLFPSSGRVVPPTYWQ 883  
DB 814 TGRSLMSVAALAGMRREQPLPHGLRGILLTELHRAAGALDYSALYPA-GRLVDAPLPAWT 872  
QY 884 RERYWID----- 890  
DB 873 HAFLFTDDGQEQRAQACTIVHPLLGSHVRLTEEPERHWOGDVGTSLWLSDHQVH 932  
QY 891 ----- 890  
DB 933 NVAALPGAAVCEMALAAAEVGEALAEVNDITPEQMLLDEQTPIDAVASIDAQGVNFT 992  
QY 891 -----TKADDAARGDRAFG----- 905  
DB 993 VETNRDGETTRHATALRAAEDDCPPGYDITALLQHPHAVGTAMRESFAERGVTLGA 1052  
QY 906 -----AGH-- 908  
DB 1053 AFGGLTTAHTPRRPRCWRSPRSCPRSSSSRAPESTTRCWTLYSSRARASRYGHW 1112  
QY 909 -----DEVEEGC--AVRG----- 919  
DB 1113 PAVAVGACAPACLRATVPCPRLHNAVDOGLQRDPWEMADLVLDHGTVLLAVRLGRNG 1172  
QY 920 -----GDRR-----SARLDHP----- 930  
DB 1173 TGTSEDERDRLVSEBRLTLGMQORALPEVGDEAGSMWLLIDTSNAVDTPDMLASTLTA 1232  
QY 931 ----- 930  
DB 1233 LKSHGQGTGECASLMSVODTPPNDOAGLEKLSQLRGDDGVIVYGVPRVGDPSHLA 1292  
QY 931 -----PPSGRRREKVEAAGDR----- 946  
DB 1293 GRQVHLVRIYTELAEFEGELPRLFVYTRQAVYVPHSGEGRANLEQAGRLRLAVIS 1352  
QY 947 ----- 946  
DB 1353 EHPMLRTLLIDVDEHTDVERVAQQLSGSEDEETAWRNGDWYARLTTPSLGHEERRTA 1412  
QY 947 -----PFLRLEIDPGLVLDLVLVYTERARAGLGEVLEIVADAAGLSFNVOQLAGVPP-- 998  
DB 1413 LBDPDHGMQRVQVRPDDLTLEFVASDRVPPGGQLEVAASMSINPADVLIHFGFPPII 1472  
QY 999 DDLPGKPNPRLILGEGCAGRIIVAVGCVGLVVGQPVIALS-AGAFATHTVTSAAVLPR 1057  
DB 1473 DD--REPQ-----LGMFVGVTYAVGSGYTGHOVGRDVGGSEBGCRTPLTCANLAVLT 1526  
QY 1058 POALSIAEAMAVAYLTAWAYLDRJARLQPGERVLIHAATGGVGLAAVQAOHGAEVH 1117  
DB 1527 PPGTLDEQATTAATTAHTATAYGLNDLAQIKAGDKVLIHSATGVGOAALISARAKAEIF 1586  
QY 1118 ATAGTPEKRAYLESIGRYVSDSRDPFADVAVAMTGEBGVNDVNLSTSELDXSFNLL 1177  
DB 1587 ATAGNPAKRAMLRDMGVHEHYDYSRVSFEAQRIRDTDGYVDIVLNSLTGAAGRAQLELL 1646

QY 1178 RSHGFVELGKRCYADNQLGRLPRLNLSFSLVDLRGMLEPRPARVALLELLGLIA 1237  
 DB 1647 PFGEFVEIGKADYVGNTRLGFPFRRLGTFEYLLDLAMSVQPRVRELLATVRLNAD 1706  
 QY 1238 GVTFPPPLATLPIAAVAADAFSMAOAHGKVL 1271  
 DB 1707 GVLTAPQCTHYPLAEMADAIRAMSNAEHTGKLV 1740

RESULT 6  
 PSB\_MYCTU STANDARD; PRT; 1538 AA.  
 ID PSB\_MYCTU  
 AC 010978; OS3234;  
 DT 01-OCT-1996 eRel. 34. Created  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Phenolphthalein synthesis polyketide synthase psb.  
 GN PSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634220;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,  
 RA Stulson J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J.F., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kojima J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPHTHALEIN  
 CC SYNTHESIS.  
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETINE (By  
 CC similarity).  
 CC -1- SIMILARITY: Contains 1 acyl carrier domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AL021070; CAA15929.1; -  
 CC EMBL; AE007122; AKA7329.1; -  
 CC PIR; E70874; E70874.  
 CC TIGR; MT3002; -  
 CC TubercuList; RV2932; -  
 CC InterPro; IPR001227; AC transferase.  
 CC InterPro; IPR000794; Ketoacyl-synt.  
 CC InterPro; IPR006163; PP\_bind.  
 CC InterPro; IPR006162; Pantne\_attach.  
 CC Pfam; PF00698; Acyl\_transf; 1.

DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt C; 1.  
 DR Pfam; PF00550; pp-binding; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; FALSE\_NEG.  
 DR PROSITE; PS50075; ACP\_DOMAIN; 1.  
 DR Hypothetical protein; Multifunctional enzyme; Oxidoreductase;  
 KM Transferase; NADP; Phosphopantetheine; Complete proteome.  
 FT DOMAIN 145..1495 ACP\_CARRIER (ACP).  
 FT BINDING 1458..1458 PHOSPHOPANTHETINE (POTENTIAL).  
 SQ SEQUENCE 1538 AA; 162527 MW; B55E2A2042AD00CC CRC64;

Query Match 22.1%; Score 1592.5; DB 1; Length 1538;  
 Best Local Similarity 31.2%; Pred. No. 2.2e-76;  
 Matches 488; Conservative 207; Mismatches 600; Indels 271; Gaps 44;

QY 3 DRPIERAEDPIAIVGASCRPGYIDISGWTLLSEGRDVTGVRPAERWDAAMFDDP 62  
 DB 25 DRRVRIAAEPPVAVGICRPPGVDGPPSEFMDPLVAGRNASTVPADRMDEAFYHDDP 84  
 QY 63 DAPGTPVTRASFSDVACPDASFPGISPRRLMDPAHRLLLRCWCALEMAIAPSL 122  
 DB 85 LTPGRMTTKGCGFPDVAAGPDAEFGITPRRAAAMDPOKMLVEALEHAGIIPPSL 144  
 QY 123 VGTETGVFIGIPSEYEALPQATASAEIDAAGLGTPSVGAGRISYALGRPCVAVD 182  
 DB 145 GGTHTAVVMGVYFNEYQSMU--AASPQVDAVSGTGNHNSITVGRISTYLLGRGPAVAVD 202  
 QY 183 TAYSSLYAVHLAQCISLRSGECSTALAGVSLMSPSTLVLSKTRALARGRCAASAE 242  
 DB 203 TACSSLYAVHLAQCISLRRETDLALAGVSITLRPEQIATISWGLTSPGRCAPDA 262  
 QY 243 ADGGRGRCGCAVVVKKLSGRADGPHLAVIRGSATINHDASSGLTPNSSQEIYK 302  
 DB 263 ADGVRREGACVVVKKLTDVNRGDQVLAVRRSAVNDGRSNGVAPNTPAKCDVAVD 322  
 QY 303 ALADAGCAASSGVGEAAGTGTGDPLEIGALAAVVGIGRDVATPLLIGSVKTNLHPE 362  
 DB 323 ALRSGDVAPDSVNVVEAHGTGTVDPIEPFALAAATGHGDACA--LGVKTNIGHLE 379  
 QY 363 YASGITGLKVLVSLQHQIPAILHQAALNFRISWGDLRLTVTRATPWPMPNTPRRGV 422  
 DB 380 AAAGIAGIKATLAVORATIPPNLHFSQWNPALDAAGTRFPVQNSWEPFAEGRRAAV 439  
 QY 423 SSFMSGNNHVVLEE-APATCTPPAPERAELLVLSARASALDQAARLRLHLEYTP 481  
 DB 440 SSFELGGTNAHVILEQSSSELAPVSEGEDTGVSTLVVTKTAQRAAATAQVLAMMEGPG 499  
 QY 482 SQ-CLGDAVAFSLATTRAMEHRLAVAAATSRGLRAALDAAGOTSPCAVRSIADSRGK 540  
 DB 500 AEVAVADVAVHNVNHRARQATFGTVVARDRAQALAGRLAAGAHAGVV-SHODGSFGR 558  
 QY 541 LAFLFTQGAQTLGMRGLYDVMSAFREAFDLCVRLFENDELRLPRLVEVMAEPASVDA 599  
 DB 559 GTVFVYSGRSGQMGMRQLADAPAAVALELPFVEQAGSLRDVI--ATGKELV 615  
 QY 600 LLDQATQPRALFFFEVALAAWRSKGVREPLVAGHSIGELVAACVAGVSELEDAVELVA 659  
 DB 616 GIEQ--IQGLIGMQTLLELWRSYGVQPLVIGHMGEVAAVAVGALLPAGRLRTA 672  
 QY 660 ARGILMOALPAGAMVIEAPADVAAVAPHAASVSIANVANPDQVVIAGAGPVAHA 719  
 DB 673 TRAILMAPLSGGGSMALLGDAATEALADIY-PQVTVGIRNSRQVYIAGPTQIDBLI 731  
 QY 720 AAMARAGRTKALHVAHSHAFSLPMAPLAEFGVAESVSYRPSIVLVSNLSGACTDEV 779  
 DB 732 ARVAAQRFPASRVNIEVAPNPAMDALQPMRSELDLTPRTPIGIIISTVYADLHTQPI 791  
 QY 780 SSPGVYVRAHREVVPFADGVKALHAGAG-----TFVEVGPXKSTLGLVAPACMPDA-RP 832  
 DB 792 FDAEHMAITNNKNNPFRFO--QATASAGDAGVHTTETISAPLLTQALADTLEDAHRP 848  
 QY 833 A-----LLASSRAGRDEPATVLEALGGLMAVGVGLVWAGLFPSSGGRVPP----- 876



QY 118 APSALVGTETGVFIGIGPESEYEAALPOATAS--AEIDAHGGLGTMPGVGAGRISYALGRLG 176  
 Db 208 PGTLRSTATGVPAAGCSEYGA---MABADLSQVDMGMSNGSAMSIIANRLSYFLDLNG 264  
 QY 177 PCVAVDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMSPBTLVWLSKTRALADGRC 236  
 Db 265 PSVAVDTACSSSLVAHLACQSLRSTQDCHALAIAGVNLSPVAFRGDQVGLSPYGC 324  
 QY 237 KAFSAEADFGREGGCAVVVLKRLSGARADGRIILAVITGSAIINHGSAGLTVPGSSQ 296  
 Db 325 RNFADATADGFRGEGGVVVLKRLTDQRDGDVLAVICGSAVNOGRSNGMLAAPPAQ 384  
 QY 297 EIVLKALADAGCAASSVGVVENAGTGTTCDPTEICQALNAVYGLRDVATPLLISVKT 356  
 Db 385 MAVLRAAYTNAGQBPSEVDYVEAHGTGTLGDPTEARALGTVGRPRDPSLLISVKT 444  
 QY 357 NUGHPYASGITGLKLVLSLOHQI PAHLHAQALNPRI SWGDLRLTVTRARTPMDMT 416  
 Db 445 NIGHTEAAGIAGFIKTIVLAVOHGQIPRNOHETAPHIPTDLMKVVDTQTEMPTGCH 504  
 QY 417 PRAGVSSFGMGSTNAHVLEEARATCTPRAPERB--ELVLSATTAALDAQAARD 475  
 Db 505 PRAGVSSFGMGSTNAHVLEEGQEVRAPEGGLSPAVSTLVAGKMQRVSATAGMLAD 564  
 QY 476 HLETPSQ--CLGDVAFSLATTRSAMEHRLAVATSREGLRAALDAAGQOTSPGAVRSIA 534  
 Db 565 WMEGPGADVALADVANTLNHHSROPKFCTVVARDDTQAIAGRALAAGQHPGVNP-A 623  
 QY 535 DSSRGK-LAFLFTGQAGQTLGWGRGLDYVWSAFREAFDL CVRLFNGLDRLPREVMAEP 593  
 Db 624 DSSPGCTVFFVYSGRSQWAGMGRQLADEPAFAVAELPEVFAVQAGFSLHDVL---- 679  
 QY 594 ASVDAALDQNTAFTQALFTFEALALALMRSCVEBELVAGHSIGLVAACVAGVSLD 653  
 Db 680 --ANGEELGIDQIOLIGMOLALTELCYGVPRDLVIGSMGVAAVAAGALTPAE 737  
 QY 654 AVFLVARGRLMQLPAGG--AMVSIADPADVAALVAAPASVSI AAVAPQVVIAGA 711  
 Db 738 GLRVATRSRLMAPLSGGQGMALLEADPTTEALINDP---QVTLGITNSRQVIYADP 794  
 QY 712 GQPVHAIAAAMARGARTALHVSHPSPMLAEMLEAFGRVAESVYRPSIVLSNLS 771  
 Db 795 TEQIDELIARVRAQKRFARVNI EVAPHPNADALQPMRSELADLTPRTIGITSTY 854  
 QY 772 GKACTDEVSSPGVWVHAREVVRFADGVALHAAGG-----TFEVGP-----KSTLLG 821  
 Db 855 ADLHTOPVDAEHMATMNRNPFVRFQ---QALASAGSGADGAYHTFEISAHPLLTQALID 911  
 QY 822 LVPACWPDARPALLLASSRAGDEPATVLEALGGLMVGGLVSWAGLP-----SGGRVP 876  
 Db 912 TLHSAQPGARYSLSGLTQDRTDQVTFRTNLKNAHTI-----HPHTHTPPPHPP 962  
 QY 877 LFTYVQREKRWIITDYADDAARGDRRAPAG-----HDEVEEGAVRGDRRSARL--DH 929  
 Db 963 IFTTMOQHRHMITTYKYPAGSVGS--APRAGTLGQHTTAVATVSAPPSHLMQARLAPDA 1020  
 QY 930 PPEEGRR---EKYBA-----GDRPRLIDEPEV 957  
 Db 1021 KYQGGHREHOVEVPAVSVLHTLITSAATELGYSALSEVRFPQPIPADRPRLIQV---V 1076  
 QY 958 LDHLVLRVTERA-----P 971  
 Db 1077 ADNRAISLASSPRACTPSDRWTRHVTACQLSSPSDSASSLNHHRANQPRERAHDLIP 1136  
 QY 972 GLGEVEIADAAGLSF-----NDVQALGMVPPDLLPKPNPRLLLGECACGRIV 1020  
 Db 1137 DIAELLAMRGIDGLPFSMTVAVSQSSMLTVAIDL--PALPBGSGTPLL---DAVHDA 1192  
 QY 1021 AVGEGVNG--LVVGGPVIALSAGAFATHTTSAL-----VLPRPQALSABEAA 1068  
 Db 1193 ALSDVADSRLLVPASTEQISLQGVVTPGRSSVTLNFTAHDDGIVDTVVAHGEVPSLS 1252

QY 1069 M-PVAVLTAMVALDRIARLOPGERVLVIAATGVC-----LAAYON-----AQHV 1112  
 Db 1253 MSLRYRLDLDFGLD--VGAQP-----PASTGPVEAYCOATNFVHTIDMQPTVPDATHP 1305  
 QY 1113 GAE--VH-----ATGTPKRAYLESIGRVYSDRSRDFVAD 1148  
 Db 1306 GAEQVTHPQPAVLIIDDGALCETLEGAGYQPAVMSDVSQARVYVYVADSDPAGADETD 1365  
 QY 1149 ---VRAMTGCSEG----- 1157  
 Db 1366 VDFPAVRLCTEITGLVTRTAEADADKPAALMILTRGVHESVAPSALRQSLKGLAGYIAE 1425  
 QY 1158 -----VDVNLNLSGELIDKSNFLRSHGRFVELGRD----- 1190  
 Db 1426 HHEMLGVLDLALINDLGEFGLPALLAKPSKSI--LVRRDQVLAALAPVGRBAPARKS 1484  
 QY 1191 --CYAD-----NOGL-----RFLRNLFSVLVD----- 1212  
 Db 1485 IQCRPDAAVYLITGGLGALGLMADWLADRGARHLVLTGRTPLPBRDMDLDTLDELRRR 1544  
 QY 1213 -----LRGMLERRPAR-----VRALL-----EELGLI----- 1235  
 Db 1545 IDAITALEKRGVTEAVALADVGCREDVQALHARDGCAPIRGIIHAAGITNDOLVISM 1604  
 QY 1236 -----AAGVFTPPP-----IATLP 1249  
 Db 1605 TGDVAVQVWMPKIGSOVLHDAFPFGSVDFYLTASAAGIFCIPGGSYAAANSYLDALA 1664  
 QY 1250 IARVADAFRSM-----AQQHGLKVLVTLGDPF----- 1277  
 Db 1665 RARROGGCTHMSLDVNAWRGLGLADAOVLSEELRMSGRDITPSEAFPAWEFVDGYDVA 1724  
 QY 1278 --VOIRIPTHAGAGSTGDRDLIDLASAPAR--AAALEAFRTQVSQVLRTEIKVG 1333  
 Db 1725 QAVVPMPPAGADSGANAAVILPARNWSVMATVRSLEEGRLRIIAELRVEKED 1784  
 QY 1334 AEALFTRLGMSIAMEVLENRIRIASLKLSTTFLSTSPNIALQNLDALATALSLE 1393  
 Db 1785 TDRPFAELGSLNLSMAVIRAEQGVGIELSTMLFNPHTVASLASYL-----AKR 1835  
 QY 1394 VAENLRAGVQNDPVSSGA 1412  
 Db 1836 VAPHDVSDNDQISALSSSA 1854  
 RESULT 8  
 WA EMENT  
 ID WA EMENT STANDARD; PRT; 1986 AA.  
 AC 003149;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Confidial Green pigment synthase (EC 2.3.1.-).  
 GN WA.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC Eutroiales; Trichocomaceae; Emericella.  
 OX NCBI\_TaxId=162425;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93101122; PubMed=1465094;  
 RA Mayorga M.E.; Timberlake W.E.;  
 RT "The developmentally regulated Aspergillus nidulans wa gene encodes a  
 RT polypeptide homologous to polyketide and fatty acid synthases";  
 RL Mol. Genet. 235:205-212(1992).  
 CC -FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN  
 CC INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED  
 CC BY CONFIDIAL LACCASE TO FORM THE GREEN PIGMENT IN NATURE  
 CC ASXULI SPORES (CONIDIA).  
 CC -CORPACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETHEINES  
 CC (POTENTIAL).  
 CC -PATHWAY: Confidial green pigment biosynthesis.  
 CC -SIMILARITY: Contains 2 acyl carrier domains.



CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X65866; CAA6695.1; --  
CC PIR: S28353; S28353.  
DR InterPro: IPR001227; Ac transferase.  
DR InterPro: IPR000794; Ketoacyl-synt.  
DR InterPro: IPR006163; Pp bind.  
DR InterPro: IPR006162; Pantane attach.  
DR InterPro: IPR001031; Thioesterase.  
DR Pfam: PF00698; Acyl\_transfer; 1.  
DR Pfam: PF00109; Ketoacyl-synt; 1.  
DR Pfam: PF02801; ketoacyl-synt C; 1.  
DR Pfam: PF00550; Pp-binding; 2.  
DR Pfam: PF00975; Thioesterase; 1.  
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 1.  
DR PROSITE: PS50075; ACP DOMAIN; 2.  
KW transferase; Phosphopantetheine; Multifunctional enzyme; Repeat.  
FT DOMAIN 529 582 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
FT DOMAIN 991 1024 ACYL/MALONYL TRANSFERASES  
FT (BY SIMILARITY).  
FT DOMAIN 1650 1719 ACYL CARRIER (ACP) 1.  
FT DOMAIN 1772 1841 ACYL CARRIER (ACP) 2.  
FT ACT\_SITE 548 548 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
FT ACT\_SITE 1001 1001 ACYL/MALONYL TRANSFERASES (BY  
FT SIMILARITY).  
FT BINDING 1682 1682 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
FT BINDING 1804 1804 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
FT SEQUENCE 1986 AA; 216634 MW; 74EF0940FF4DE9A CRC64;  
SQ  
Query Match 16.4%; Score 1180; DB 1; Length 1986;  
Best Local Similarity 26.9%; Pred. No. 1.6e-54;  
Matches 419; Conservative 24; Mismatches 65; Indels 264; Gaps 53;  
QY 8 RAADPIAIVGASCRIPGVIVDLSGFWTLGSRDVTGVPAERMDAAWPDPPDPAQK 67  
DB 373 RAELSKAIITIGMGRFP-EADSPQDFWNLLYKGLDVRKVPEDRMDADAVDLTGATNT 431  
QY 68 TPTTRASFSDVACFDASFGISPRELRMDPAHLLBVCWALEMAALAPBALVTE- 126  
DB 432 SKVPYGCWIREPGLPDRFFNMSPREALQADPAQRLALLTAYALEGAGFVDPSTBQTOR 491  
QY 127 --TGVFIGIPSEYEALPQATASAEIDAHGGLGTPMVGAGRIYALGLRGCVAVDTA 184  
DB 492 DRGIFGYMTSDYR-----EVNSGQDIDTYFICGNRAFPTRGRINITYFKSGSVSDTA 547  
QY 185 YSSSLVAVHLACOSLNSGECSTALAGVSLMLSPSTLWLSKTRALARDRCFAFSAD 244  
DB 548 CSSLAIHILACNSIMWNCDDTATITGGVILITWPDNAGLDRCGHLSRTGNCNTFPDGD 607  
QY 245 GFRGEGCAVVLKRLSGARADGRLTAVIRGSAIINHDASSGLTVNNGSSQELVLRAL 304  
DB 608 GYCRADGVTVLRLLEDALADNDPILGVINGATNNSAEVSTTRHVGAGQAFIEFKLL 667  
QY 305 ADGCAASVGVYEAHGTGTLGDIPIEIQALNAYVGG--RDVATPLLISVNTNLGHP 362  
DB 668 NEANVPKNI SYIEHMTGTQAGDAVMQSLVDFADHRRGPGQSHLSAKSNIGHGE 727  
QY 363 YASGITGLLVLSLOHGQIPAHLLAQA-INPRIWGDRLVTYTRATPMPDWTNP- 417  
DB 728 SASGVSTLVKVLMMKEMNIPHCIGIKTKINHNPPTDLAQRNVAIALQP-TAMNRPFGK 786  
QY 418 RRAQVSEFGMSGTRAHVLEAPAACTCPAPF-----RPAELVLKARTASAL- 469  
DB 787 RQIFLNFSAAAGNTALLLEDGVSD-----PEGEDKRRTHVITLSARSOTALLONNIDAL 841

QY 470 AARLRDHEETY---PSQCLGDVAFSLATTRSAMENRLAAVATSRREGRAALDAAAGQ-- 524  
DB 842 CQYISQKQETFGVKQSNALPSLAYTTARIRHHPFVTAIGSSFGQMRDLSLASSKERY 901  
QY 525 ---TSFGVRSIADSSRCKALFLFTGGCAQTLGMRGILYDWSAREAFDL--CYRLRN 578  
DB 902 AVPAKTPG-----IGFLFTGGCAQVAAAMKQLEYECSHRSRYSIEHLDICSG-G 948  
QY 579 QELDRPLREVMNAEPASVDALLDQAFQPALFTPEYVLAALMRSMGVEPELVAGHSIG 638  
DB 949 QDPLSTPLVDQSLPSEISPPVYQIGTT-----CYQALSSPFWSLGTTPSVLVGHSIG 1003  
QY 639 ELVAACVAGVFSLEDAVFLVAARGLM-QALPAGAWVSIEAPADVAANAAPHAASVS 696  
DB 1004 DPAANMAAGVLTSDITVACGRRAQQLTERCQPGTHAMLAIKAPLVEVQQL--NEKVAD 1061  
QY 697 IAAVNAPODVITAGAGQPVHAIAAAMAAGARTKALHVSHTRSPPLMAMLNFGVVAAS 756  
DB 1062 MACINSPESTVLSGPKSSIDELSRACSEKGLSTLTIVYAFRHSAGVEPILIEDLEKALQG 1121  
QY 757 VSYRPSIVLNSLGGKACTD---EVSPPGYWRHAREVYVRFADGVKALHAAGAG---T 809  
DB 1122 ITFNKPSVPPVSALLGEVITTEAGSNILMAEYLVRHRETVNPLSAREAVRNKLGSDQTL 1181  
QY 810 FVEVGKSTLGLVPACMDAPALLASSRAGREDEPATVLEALGGLMAVAGLVSMAGL-- 867  
DB 1182 WIEVGHTVCSGVNKA TL-GPOTTWASLRBEDTKVLSNLSLYLAGVDINWQYHQ 1240  
QY 868 -PFGSGRRVPLPTYPQRRRYWID-----TKADDAAGDRAPAGHDEVEGGAIVG 919  
DB 1241 DFSSSHRVPLPTKYKDLKNYMI PYRNNECLTKGSSMS-----AASASLOPTFLTTSQNRV 1296  
QY 920 GDRSARL-----DHPPESSGR---REKYEAGDRPRLIDEPPVLDHLVATRTR 968  
DB 1297 VESRDGLTATVYVNDIADPDLNRVIQGHKNAGALCCSSLYADBAQTLAEVLI--EK 1353  
QY 969 RABGLGEVEIADVADAGLSFNDVQALAGVPPDLPGKPNPPLLIGECACRIYAVGEGVNG 1028  
DB 1354 YKREL-----KGSGLDVCNVTV-----PKPLIAKTGKEQFRISATAMVWDK 1394  
QY 1029 LVVGQFVIALSAG-----AFHTVITTSALVLP RPALALAEIAAMP--- 1070  
DB 1395 HVSQVFSVYAEKKLIDAHCEVKLPDCMAADLEWKRSYLVK-RSIELLENSAVKQDA 1453  
QY 1071 -----VAYLTAVYALDIARLOPGERVLI-----HAATGVGGLAAVQ-----WAO-- 1110  
DB 1454 HRLRGVYKLSBALVDYDENTQSIREVILDSHEHATLALVKFQAQAMFHNRYAIDSE 1513  
QY 1111 -HV-GAEVHATACTPEKRAVLSLGVRYVDSRS--DRFVADV-----RAWT 1153  
DB 1514 GHLSGFIMNASDQTDSK-----SQVFVNHGMDSWRCLKKFSADVTYRTYVRMQPWRDSIWA 1569  
QY 1154 GG-----EGVDV-----LNSLSGELIDKSF-----NLARSHG--RVELGKRD 1190  
DB 1570 GNVYIEGDDITIAVFGVVFQALSRKILDIAPPAQLSRAQTSPIQSSAPQKIEFAKPT 1629  
QY 1191 CYADNOLGLRPFRLNLSFSLVDLRGMLEPARVRLALBELLGLLIAAG-----VFTPPPI 1245  
DB 1630 SRPAPVYTKSFYKSGAGSVVYRAL-----NILASVGLSSDSMDLVFADYGV 1680  
QY 1246 ATLPIARVADAFESMAOAGLKVLTLDPEVOIRIPF-----HAGAGPST 1292  
DB 1681 DLSLSLTVGKYRE-----ELNDM--DSVFIHPTVGDKRFVTOQSPSVASDSSS 1731  
QY 1293 GDR-----DLDDLAS-AAPAAAMAALEALTRQVOQVLR-----TPETKVGAEAL 1337  
DB 1732 TDRESEYSPNGSCSGSLSPASPGTVSPNEKVIQIHENGTMKEIRAILIADEIGVADAI 1791  
QY 1338 -----FTRLGMDSLMAVELRNRIEASIKLKLSTTFSTSPNIALAQNLLDALATLSTSE 1392  
DB 1792 KSDENINELGMDLSLTLVIGKIRBSLMDL-----PGEFFINQTLQDQETLADLDK 1843

```

RESULT 9
MSAS PENPA STANDARD; PRT: 1774 AA.
AC P22367;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DB 6-methylsalicylic acid synthase (EC 2.3.1.165) (6-MSAS).
OC Penicillium patulum (Penicillium griseofulvum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxId=5078;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=DSM 62862;
RX MEDLINE=91006137; PubMed=2209605;
RA Beck U., Rippka S., Siegmeyer A., Schiltz E., Schweizer E.;
RT "The multifunctional 6-methylsalicylic acid synthase gene of
RT Penicillium patulum. Its gene structure relative to that of other
RT polyketide synthases."
RL Eur. J. Biochem. 192:487-498(1990).
CC -!- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC It catalyzes a total of 11 steps by seven different component
CC enzymes, in the biosynthesis of the antibiotic patulin.
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH = 6-
CC methylsalicylate + 4 CoA + 3 CO(2) + NADP(+).
CC -!- PATHWAY: Patulin biosynthesis.
CC -!- SUBUNIT: HOMOMULTIMER.
CC -!- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC -!- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC THIOLEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X55776; CAA39295.1; -.
DR PIR, S13178; S13178.
DR InterPro: IPR001227; AC transferase.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR006163; Pp bind.
DR InterPro: IPR006162; Ppantne_attach.
DR Pfam: PF00698; Acyl_transf; 1.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt C; 1.
DR Pfam: PF00550; pp-binding; 1.
DR PROSITE: PS00012; PHOSPHOPANTHERINE; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS50075; ACP DOMAIN; 1.
KW Multifunctional enzyme; Transferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine.
FT DOMAIN 166 238 ACYLTRANSFERASE (ATP).
FT DOMAIN 642 676 ACETYL/MALONYL TRANSFERASES.
FT DOMAIN 1403 1450 2-OXOACYL REDUCTASE.
FT DOMAIN 1700 1769 ACYL CARRIER (ACP).
FT NP_BIND 1419 1424 NADP (POTENTIAL).
FT ACT_SITE 204 204 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 653 653 MALONYLTRANSFERASE (BY SIMILARITY).
FT BINDING 1732 1732 PHOSPHOPANTHERINE (BY SIMILARITY).
SQ SEQUENCE 1774 AA; 190732 MW; 05ED5DD10863F938 CRC64;
Query Match 16.3%; Score 1177; DB 1; Length 1774;
Best Local Similarity 25.6%; Pred. No. 2e-54;
Matches 458; Conservative 210; Mismatches 655; Indels 466; Gaps 56;
QY 14 IAIIVASGRLPGGVIDLGSFWTLBGSRTDVGVRVPERMDAAAMPDPDPAEGKTPTVTA 73
DB 35 VAVVGNACRVAGNHNHPELLMOSLLSOKSAMGEIIPMRWEPPYRRDARNEKFLKNTTSRG 94

```

```

QY 74 SFLSDVACEDASFFGISPREALRMDPAHRLLEVCMEALENNAIAPSAVLGTETVFIGI 133
DB 95 YFLDLLEDPDQCFPGISPREAQNMPQGRVSLVSAENLEDGDIKAKSLSGSDTAVFMVY 154
QY 134 GPSEYEA---LPQATSAEIDAHGIGCTMPSVAGRISYALGRPCVAVDTAVSSGL 189
DB 155 NSDDPSKVLVDLP-----NVEAMWGGITACVGPNNRISYHNLNMGSTAVDAACASSL 208
QY 190 VAVHLAQSRLRGCESTALAGVSLMLSPSTLVLSKTRALARDRCRKFSAEADQFGMG 249
DB 209 VAIHHGVQAIRLGESEKVALVGVNALCGGLTRVLDKGAISDSCSFFDDAHGYAAG 268
QY 250 ECGAVVILKRLSGADGDRILAVIRGSAINHDSAGSGLTVPNGSSOEIVLKRALADAC 309
DB 269 EAGAGLVLSKRLALDHIDNVAVIKGSAVCDGDKTNGIMARNNSVAQOLAAANNLSAANI 328
QY 310 AASVGYEAGGTGTTLDGPIEIOALNAVYGLGRDVAIPLIGSVKTNIGPEVASGITG 369
DB 329 DHTVRVYEAHATSTPLGDPTEISAIASVYGAADRPADPCYIGSIKPNIGHLEAGGVWG 388
QY 370 LKAVTLISQHQIPAHHAQALNPRISMGDLRLTYTRARTPMDNTPRRAGVSSRGMG 429
DB 389 FIKAVLAIQKGVLPQANLTKLNSRI DMKTAGVKVQEAETPWPESDPIRRAGVCSYGGG 448
QY 430 TNAHVLEE-APAATCTP---PAPERPAELVLSARTASALDAQAARLDRHLEY-PSQC 484
DB 449 TVSHAVIEEFPILDPDPLGNAGVSGPG-LILSSPQEGRLALQAKTLRDMWTAGCKDHN 507
QY 485 LGDVAFSLATTSAMEHRLVAATSREGIRALDAQAQ--OTSPOAVRSIADSSRGKLA 542
DB 508 LSDLITLTLATRDHHDVRAALVDDYDRDAEQVLSGLANGVDHTFTQSRVLSDISKDV 567
QY 543 FLFTQAGQAGTGMGGLVDWMSAFREAPDLCLRLNQEIDRLREVMAREPSVDAALLD 602
DB 566 WFSFGHQWPMWQKOL-----IHNVPFAAL-----QPLDEIQEIGLSPLELR 614
QY 603 QTAPE-----TOPALTFEYALALRSMGVSEBELVAGHSIGELVACVAGVSLDEAVPL 657
DB 615 TDPFSSDRVQVLTLYVMQIGLSALLQNSGITQVAIGHSVGSIASVAGALSAPBAGALI 674
QY 658 VAARGLMOALPAGAMVSIAPADVAANAAPHASVIAAVNAPDOVINGACOPVHA 717
DB 675 YRRRLALYRQWKGKMLVNPISAETETILSRSRLV-VAIDSSPSCVAVAKDELVAE 733
QY 718 IAAAMAAGARTKALHVSIAFHSPLMAPLAEFGSV-DESVYRRPSITYLVNLSGKACT 776
DB 734 TAEALKAGVKTFTVKSIDIAFHSPTLNGLVDPDRVLAETLSPVSPNVLYSTALADPRG 793
QY 777 DEVSSPGVVRHAREVRFADGVKALHAAAGTFFVEGPKSTLLGLVPACMPDAR----- 831
DB 794 QDLRDEYVAGMNVNRKLTSAVKAAVEDGYRLFLEVSHFPVSHSINETLADHAGMEDRA 853
QY 832 --PALLASSRAGRDEPAT--VLEALGGLMAGVLVSMGLFPSSGRARVPLPTYPQOREY 887
DB 854 VIPTLL-----KKVPTKHLHLSIAQLHCRGAENVMAAQMP-GMWATCVPTTWMHKDI 906
QY 888 WIDTADDAARDRRAP---GAGHDEVEEGG-----AVRGDR--RSARLDH---PRPE 933
DB 907 W-----RKIETAPLHTGLTHD-VEKHLLGGRIPVPGDTVYVYTRLDNDTKRPFPG 956
QY 934 S---GRREKVEAAG-----DRPFRLEI----- 952
DB 957 SHPLHGTETVPAAGLINTFLKGTGOMLONVLRVPVAINAPRSVQVVVQDQVAVSLR 1016
QY 953 --DEPGVDHLVLRVT-----ERRAPGLGE-----VEIAV 980
DB 1017 ISEPSQDLDDDSVTHTTAYVDRKAVGSEDRIDPAVKSRLVTKLADNFSIDYLDKVCV 1076
QY 981 DAAAGISF-----NDVQALGM-----VPDLP-----GK---PRP 1007
DB 1077 SAMGPFMAVTEHYRNDKEMLARVDVNPALISGAPLPMDSSSWAPVYLDAAATSVSGTTFPPT 1136

```

QY 1008 PLLL-----GGECAIRIVAVG----- 1023  
 DB 1137 ALMPQIERVEVFTSDPPKIMLVQEAQSDSPVSHSVSEAGVLAQKFMNPFSEI 1196  
 QY 1024 EGVNGL-----VVGQPVIALSAGAFTHVTTAALVLPPOALS-AIEAAMP----- 1070  
 DB 1197 EGTGVGSGMESLVHQ--IAMPATPAPPEPLSIEIVLVSPDATTTRALVAPASPTRVNPF 1254  
 QY 1071-----VAAYL-----AMVADIRIARLOPE-- 1090  
 DB 1255 QESSTQEFPSNASSLPLEKGTVTYTIPEVAISLAEPVPAESEFTWMLLEIKTQVNSLP 1314  
 QY 1091-RVLIHAATGVGVLAAVQMAQ-----HYGAEVHATAGT-----PERKAYLESL 1132  
 DB 1315 IKFPTLTANIGEOPTPALAQSPVGLARVASEHPDGLTIDVEBPVPLSTMRYIQGA 1374  
 QY 1133 GVRVSD--SRSDRPVADV-----AMTQSGCV----- 1158  
 DB 1375 DIRINDGARTSRFSRLPNKLLPASEGPRLLPREGTYLITGGGLGLEVADFLVEK 1434  
 QY 1159-----DVVNSLSGELIDKSFNLSHGRFV-----ELGKRCQYAD 1194  
 DB 1435 GARRLLISRALLPRTTQVSEEDLOPTI--AKIRLESRGASVHVLPIDITKPDVVEQ 1492  
 QY 1195 NQGLRPFRLNLSF--SLVDRG-----NMLESPPARVALLLEELGLIAGVFTPP 1243  
 DB 1493-----LTTALDRLLSPVQGVVHAAGVLDNELVWQTTDRANRVLAIPALALHVEFPF 1548  
 QY 1244 P-----IATLPI--ARVADAF----- 1257  
 DB 1549 KSVDFPVMFSSCGNLVFTQASVSGNAFLDTLATHRARLGDAAVSPQWTSWRGLMGGA 1608  
 QY 1258-----RSMAGQHLGKL-----VLTGDPF 1277  
 DB 1609 STDFINAELESKGTITVTRDEAPRAWOHLAKYMDHCVLRSAFEDGEPIPVISINDIA 1668  
 QY 1278 VQIRIPTHAGAPSTGRDULDLRLASAPARAALAEPLRTQVSYQLRTPPEIKVGAEL 1337  
 DB 1669 VR-RVGIVSNTSPAAAGSS--DAVPTSGPELK-AVDEKIRGCYAKVLCQMTAEVDVOSKAA 1724  
 QY 1338 FTRLGNDLSMAVLRNRIRIEASLKUKUSTTSLTSPNIALIAQNLIDALA 1386  
 DB 1725 LADLGVDSTVTLRLRQLTLKIAVPPTLWSPHTLWFAEKL 1773  
 RESULT 10  
 PKSI\_ASPPA STANDARD; PRT; 2109 AA.  
 AC Q12053;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Aflatoxin biosynthesis polypeptide synthase (PKS).  
 GN PKSL1.  
 OS Aspergillus parasiticus.  
 OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OC NCBI\_TaxID=5067;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=NRRL 2999;  
 RX MEDLINE=96042102; Pubmed=7592391;  
 RA Feng G.H., Leonard T.J.;  
 RT "Characterization of the polyketide synthase gene (pksl1) required  
 for aflatoxin biosynthesis in Aspergillus parasiticus.";  
 RL J. Bacteriol. 177:6246-6254 (1995).  
 CC -1- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF  
 AFLATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.  
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE (By  
 similarity).  
 CC -1- PATHWAY: Aflatoxin biosynthesis, first step.  
 CC -1- SIMILARITY: Contains 1 acyl carrier domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.embnet.org/licenses/embnet.html> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 DR EMBL: LA2766; AAC1675.1; -  
 DR EMBL: LA2765; AAC1674.1; -  
 DR PIR: T17490; T17490.  
 DR InterPro: IPR001227; Ac. transferase.  
 DR InterPro: IPR000794; Ketoacyl-synt.  
 DR InterPro: IPR006163; Pp. bind.  
 DR InterPro: IPR006162; Pplant. attach.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam: PF00698; Acyl\_transf. 1.  
 DR Pfam: PF00109; ketoacyl-synt. 1.  
 DR Pfam: PF02801; ketoacyl-synt. C; 1.  
 DR Pfam: PF00550; Pp-binding. 1.  
 DR Pfam: PF00975; Thioesterase; 1.  
 DR PROSITE: PS00075; ACP DOMAIN; 1.  
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; FALSE\_NEG.  
 DR TRANSFERASE; Acyltransferase; Phosphopantetheine;  
 KW Multifunctional enzyme.  
 FT DOMAIN 374 805 BETA-KETOACYL SYNTHASE.  
 FT DOMAIN 1714 1745 ACYL/MALONYL TRANSFERASES.  
 FT DOMAIN 543 543 ACYL CARRIER (ACP).  
 FT ACT\_SITE 543 543 THIOESTERASE.  
 FT ACT\_SITE 993 993 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
 FT ACT\_SITE 993 993 ACYL/MALONYL TRANSFERASES (BY SIMILARITY).  
 FT BINDING 1746 1746 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT ACT\_SITE 1937 1937 THIOESTERASE (BY SIMILARITY).  
 SQ SEQUENCE 2109 AA; 230715 MW; CB701372A1608551 CRC64;  
 Query Match 16.3%; Score 1173.5; DB 1; Length 2109;  
 Best Local Similarity 27.2%; Pred. No. 3,8e-54;  
 Matches 409; Conservative 222; Mismatches 625; Indels 247; Gaps 54;  
 QY 9 AEDPPIAIVGASGRLEPGVYIDLSGFVTLLEGSDTYGRPAEHWDAAMFDDPDPAFGKT 68  
 DB 369 AGCKXLAIVMSGRFPESEPTTES-FWDLLYKGLDVCKEVRRRWDINTHVDPGKARNG 427  
 QY 69 PVTRASFLDVACFPDASFPGISPREALRMDPAHRLLEVCMEALENAALAPSALVGTG-- 126  
 DB 428 ATKWGCWLPDSGDFRFFGISPKKAPQMDPQRMALMSTYEMERAGLVPTTSTQND 487  
 QY 127 -TGVFTIGPSEYEALPQATASAEIDAH---GGLGTMPVAGRISYALGLRGCYAV 181  
 DB 488 RIGVFHGVTSNDW---METNTAQNIDTYFITGNGRFP---GRINCFEPAGPSYTN 539  
 QY 182 DTVYSSSLVAVHLACQSLASGECSTLACGVSIMLSPTLWMLSKTRALAPDRCAFA 241  
 DB 540 DTVACSSSLAHLACSLKWRGDDTVAGCTNMIYPPDDHTGIDGKFFLSRTGNCPCPYD 599  
 QY 242 EADFGRGSCAAVVLKRTSGARADGRILAVRGSAINHDGASGLTVPNGSGQRIVLK 301  
 DB 600 KADGICRAESGVTFPKRLLEDALADNDPILGVILDAKTHNSAMSEMTPRHVAQIDNMT 659  
 QY 302 RALADAGCAASVGVYEAHGTGTTGDPLEIQALNAVY--LGRDVAFTPLIGVYKTNL 358  
 DB 660 AALNTTGLHPNDPSYIEHGTGTQVDAVEMESVLSVFAPEYARADPLFVGSAAKANY 719  
 QY 359 GHEPIVSGITGLKLVLSLQHQIIPHL--HAQALNPI-SNGDRLVTVTRARTWPPW 414  
 DB 720 GHEGEGVSGTSLIKVLMQHQDIPPHCGIKPKSKINRFPDLAGRVNVAIAFEKPKWPT 779  
 QY 415 NPEPRAGVSPFGSGTNAVVLLEAPAACTPPAPERPELVLSARTASALDAQARLR 474  
 DB 780 HTRRVLLNNFSAAGNTALIVDAPERMPTKDRSSHIVALSAGVASKMTINERLH 839  
 QY 475 DHLETYPSOCLGDVAFSLATTRSNMHRLAVATSRREGRLAALDAQAQ--GQTSFGAVR 531

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:33 ; Search time 49.4261 Seconds  
(without alignments)

2764.850 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210

Sequence: 1 VADRIERAEDPIAIVGAS.....GVQNDVSSGADQDWETIAL 1421

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2331	32.2	4151	2	G70944
2	2298.5	31.9	2103	2	G86825
3	2211.5	30.7	2478	2	AH2140
4	2207	30.6	1502	2	H70984
5	2136	29.6	2518	2	A12140
6	2120	29.4	1587	2	AB2012
7	2093.5	29.0	3573	2	S23070
8	2063.5	28.6	3413	2	T17467
9	2052.5	28.5	3739	2	T17410
10	2051.5	28.5	4613	2	T17409
11	2025	28.1	2126	2	H70621
12	2014.5	27.9	2188	2	A70984
13	2013.5	27.9	6260	2	T30228
14	2003.5	27.8	4735	2	T17463
15	2000.5	27.7	10223	2	T30225
16	1983.5	27.5	2201	2	S73014
17	1976	27.4	8563	2	T30226
18	1965	27.3	6420	2	T30283
19	1955	27.1	2100	2	T03223
20	1925.5	26.7	3491	2	T43231
21	1908.5	26.5	2108	2	H70819
22	1900.5	26.4	2124	2	T28658
23	1889	26.2	3519	2	S43048
24	1875	26.0	3172	2	S22012
25	1869	25.9	2126	2	E70522
26	1855	25.7	2111	2	A70668
27	1852.5	25.7	2110	2	B44110
28	1831.5	25.4	5069	2	T17464
29	1827	25.3	2116	2	C86926

30	1812	25.1	1774	2	T17421	polyketide synthase
31	1809	25.1	7576	2	T17428	FK506 polyketide synthase
32	1800	25.0	3178	2	S13595	6-deoxyerythronolide synthase
33	1798	24.9	1562	2	T17411	polyketide synthase
34	1790.5	24.8	1402	2	D70634	probable polyketide synthase
35	1785.5	24.8	1763	2	T17465	rifamycin polyketide synthase
36	1784	24.7	1570	2	AC2012	hypothetical protein
37	1776.5	24.6	1762	2	T03222	probable polyketide synthase
38	1773	24.6	1728	2	T17466	rifamycin polyketide synthase
39	1751	24.3	1937	2	T03224	probable polyketide synthase
40	1719.5	23.8	1017	2	B70985	probable polyketide synthase
41	1712.5	23.8	2118	2	S72705	mycrocetate synthase
42	1684.5	23.4	1293	2	T30871	orellin acid synthase
43	1678	23.3	2297	2	T34918	polyketide synthase
44	1643.5	22.8	1827	2	B70984	probable polyketide synthase
45	1638	22.7	1346	2	T17412	polyketide synthase

## ALIGNMENTS

## RESULT 1

G70944  
Probable polyketide synthase Rv2048c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: G70944

R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentile, S.; Hamlin, N.; Holtrold, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G. A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70944

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-4151 <COL>

A:Cross-references: GB:AL021899; GB:AL123456; NID:93242282; PIDN:CA17262.1; PID:9286678

A:Experimental source: strain H37RV

A:Gene: pks12

C:Genetics:

C:Superfamily: Mycobacterium tuberculosis probable polyketide synthase Rv2048c; 3-oxoacyl homology; [acyl-carrier-protein] S-malonyltransferase homology

C:Keywords: carrier protein

F:55-45/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <ONS>

F:559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

F:1680-1861/Domain: short-chain alcohol dehydrogenase homology <SAD1>

F:1963-2038/Domain: acyl carrier protein homology <ACP1>

F:2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <ONS1>

F:2582-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

F:3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>

F:3995-4066/Domain: acyl carrier protein homology <ACP2>

Query Match 32.2%; Score 2321; DB 2; Length 4151;

Best local similarity 33.2%; Pred. No. 6; 7e-117;

Matches 613; Conservative 203; Mismatches 535; Indels 498; Gaps 31;

QY	6	IRRAEDPIAIVGASCLPGYIDSGFWTLIESGRDVGKPAER-WDAAMFDDPDA	64
DB	2051	YRTTEDPIAIVGACRYFGVNSPDDMDMIDQGRDVISEPPADRGMDLQVNDPDA	2110
QY	65	PKRTVTASPLSDVACDASFGISPREALMDPAHLLFVCEALENAIAPSAVUG	124
DB	2111	AACTTRGGFVDGDPDPAFVGPSBALMDQHRNLELSWEALERAIDPGLG	2170
QY	125	TETGVFIGISPEVEALPQATASAEIDAAGGLGTPMSVAGRIYSALGRPCVAVDTA	184
DB	2171	SATGVFAGVMQGY-----GMAAPVGVFRITGLSSVASRVAIVYGLSPAVSVDTA	2225
QY	185	VSSSLVAHVLACQSLRSGECSTALAGVSLMSBSTVLWLSLTRLLARDGRKASAEAD	244
DB	2226	CSSSLVAHMAVGSILRSGCDLALAGVTVNATPDI FVEFSRRLGSLSPDGRCKAFAMAAD	2285

QY 245 GFGREGCAVYVLRKLSGARADGRILAVIRGSAINHDSAGSLTVPNGSOEIVLKRAL 304  
 DB 2286 GTFPSBGGKMLVILORLSDARRLGHPIVLAVVGSANVDGASNGITLAPNGSQQCVVNAAL 2345  
 QY 305 ADAGCAASVGVYEAHGTGTTGLDPIEIQALNMYVGLGR-DVATPLLIGSVKTNLGHPEY 363  
 DB 2346 ANGLSAAEVDVVEGHTGTTGLDPIEIQALNMYVGLGR-DVATPLLIGSVKTNLGHPEY 2405  
 QY 364 ASGITGLKLVLSLQHQITPAHLHAQNLNRIKSGDLRLTVTRAKRTWPMPTPRRAGVS 423  
 DB 2406 AAGVAGYIKVYLMRHEHLPLTLHVDPVSPHVMASAGAVELLTAFRVWPGARTRRAGVS 2465  
 QY 424 SFGMSGNNAHVLEEAATCTPAPERPAELLVLSARTASALDAQAARLDHLETPSQ 483  
 DB 2466 SFGISGNNHVITLAVDVVRRREGAGMPVPMVVSXSSSALRGCAHRLAAYVRGDDG 2525  
 QY 484 CLGDVAFSLATTSAMEHRLAVATSRGRLAALDAAGQTSFGAVRSIADSSRGRLAF 543  
 DB 2526 DVADVGSILA-GRSVFEHRAVYVGGDRDLAAGLDELADQDGGSVVRGTA-TRAGKTVE 2583  
 QY 544 LFTGCGAQITLGMGRGLYDVMSAREAFDLCYRLFNQGLDPLLEVMMAEPAASVDALLDQ 603  
 DB 2584 VFGQGSQWLGKMGGLHAGYVPFAEAFTVVGELDRHLRLPREVMGH-DEMLNS 2639  
 QY 604 TATFOPALFTFEYALALNRSWGVPELVAGHSIGELVAAQVAFSLDPAVFLVARGR 663  
 DB 2640 TEPAQPLFAVEVALFRLGSKGVPRFVNGHSIGELSAHVAQVSLERNAALVVAARGR 2699  
 QY 664 LMOALPAGAMVSIIEADVAANAAPHASVSIAVANADQVVIAGAGOPVHAIAAMA 723  
 DB 2700 LMOALPAGAMVVAQAAEEVERPL-SAEVDIAANVGPASLYISQAQNAVAADVADQLR 2756  
 QY 724 ARGARATLAVSHAFHSPLMAEAFGRVASEVSTRPSIVLVSNISGACTDEVSSPG 783  
 DB 2757 ADGRRVHQLAVSHAFHSPLMDPMIDEFAVAAGALIGRPTIGVISNTGGLADGDFSSA 2816  
 QY 784 YWVRHAREVVRPAAGVVALHAAGTVEVEGPKSTLLGLPACMPDARPALALSSRAGR 843  
 DB 2817 YWVRHAREVVRPAAGVVALHAAGTVEVEGPKSTLLGLPACMPDARPALALSSRAGR 2876  
 QY 844 EPATVLEALGGLMAVGLVSMAGLFPSPGRRVPLPTYPMQRRERYM--DTRKADAA--- 897  
 DB 2877 EPATVLEALGGLMAVGLVSMAGLFPSPGRRVPLPTYPMQRRERYM--DTRKADAA--- 2935  
 QY 898 -----RGDRRAPAGH-----DE 910  
 DB 2936 AASEHALLGAVIDLPAASGVVLTGRLSPSVQWMLADHSVAGVTIFPGAQFVELAIRAGDE 2995  
 QY 911 V-----EEG-----GAV 917  
 DB 2996 VGGGVVDESTLAAPLVLPAGSVAOVVNVNPGDESGVSVSRGVDGTGWLHAEGAL 3055  
 QY 918 RGG----- 920  
 DB 3056 RAGSABPTADLAMPAPGAVPVEVADGYQQLAERGVGYGPAFGLTAMMRGRGEVFAEVA 3115  
 QY 921 -----DR----- 922  
 DB 3176 ARIAPVGPASVSIELADGLPVLVSVMALARPVTDQQLRAAVSSSGPDLFEVTVMSPOP 3235  
 QY 923 ----- 922  
 DB 3236 SAAVEPLPVCAWGTEDDSAAVFEVSVPLAGDVVAGVYAATSSVLDVLQSWLTRDAGVLV 3295  
 QY 923 -----RSARLDHP-----P 931  
 DB 3296 VMTRGAVALLPGEDEVTLAAGAAVWGLVRSQTEHGRITLVVDSAPFLDSDSALAAVVTGEP 3355

QY 932 PESGREKVEAA-----GDRPFLEIDEFGVLDHLVLR-VTERRAE-GLG 974  
 DB 3356 QVIMRGEVYTYTARVHOSRAVGGVLLVPSDRPMRLAMSTAGTENLELIPADADLGG 3415  
 QY 975 EYEIAVDAAGLSFNDVOLAGVVPDDLPGKNPPLLGGECAGRIYAVEGNGVLVGGP 1034  
 DB 3416 QVAVASAIANRDMVIALGLYPP-----PDAMVGEACGVVLETLSNKSFAVGD 3468  
 QY 1035 VIALSAGAPATVHTSAALVLRPPQALSAIEAAMVAVALTMYALDRITARLQGERVLI 1094  
 DB 3469 VMGLPFEGGTVASTDQRLLVKVPAGWSHTAALTTTSVFATIAHYALVDLAARSGGVLI 3528  
 QY 1095 HAATGVGLAAVQMAQVGAEEVATAGTPEKRAYLESGLV--RYVDSRSRDEVADVRAM 1152  
 DB 3529 HNGTGVGMAAVALAHNLGLIEVPATL-SKGMKDTLRAMFPDDHISDSLEEDERFRA 3587  
 QY 1153 TGGEVDVVLNLSIGELIKSFNLSHGRFVELGKRDYADNQLGR-PFLRNLSFSLV 1211  
 DB 3588 TGGRGFDVVLDSLAGFVDSLRVLVAPGVFLEMGRKTDIRDPGVIAQQVPGVRYRAFDLF 3647  
 QY 1212 DLRGWMLERPARVALLLEELGLIAGVFTPPILATPIARVADAFRSMQAQHLGLV- 1270  
 DB 3648 E-----PGEDRIAQLLAEIATLFGDGLRPLDPVITFDVRCAPALRYISQARHTKVM 3701  
 QY 1271 -----LTLGDPEVOIRIPTHGA-----GP-STGDRDLDRLASA 1304  
 DB 3702 LMGWSAAGTVLITGTGMAGSAVBARHGVRLVLSRRGPPAPGAELVALLAA 3761  
 QY 1305 APARAALAEFLRTVOVSOYLTPTEIK-----VGA-EALFTRLGMSLMAVELRN 1353  
 DB 3762 GAQGVVACDAADRALLAVIADIPVQHPLSGVIHTAGALDAVWMSLTPDRDVV-LRS 3820  
 QY 1354 RIEASLKLSTFTSTSPNIAL-----LAQNLDPALA 1386  
 DB 3821 KVDAAHMLHELTLDVLSAFVWFSMAGLVGSSGQANVYAANSFLDALA 3869

RESULT 2  
 G86925  
 Probable polyketide synthase (imported) - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #ext\_change 24-May-2001  
 C:Accession: G86925  
 R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hearn, M.A.; Rutherford, K.M.  
 Nature 409, 1007-1011, 2001  
 A: Authors: Rutter, S.; Seeger, K.; Skellern, M.; Skellern, J.; Squares, R.; S.  
 A: Title: Massive gene decay in the leprosy bacillus.  
 A: Reference number: A86909; PMID:21128732; PMID:11234002  
 A: Accession: G86925  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-2103 <STC>  
 A: Cross-references: GB:AL450380; NID:g13092513; PIDN:CAC29643.1; GSPDB:GN00147  
 A: Gene: M0135  
 C: Superfamily: mycocerosic acid synthase; 3-oxoacyl-(acyl-carrier-protein) synthase I h  
 nase homology; lacyl-carrier-protein S-malonyltransferase homology  
 C: Keywords: carrier protein

Query Match 31.9%; Score 2298.5; DB 2; Length 2103;  
 Best Local Similarity 32.9%; Pred. No. 4.5e-116;  
 Matches 627; Conservative 204; Mismatches 536; Indels 541; Gaps 38;

QY 4 RPIERAADPIALVAGSGLPGGVLDLSGFMTLLSGSRDTGVRPAER-WDAAMFDPDP 62  
 DB 34 REVEQDAPTEPVAVVIGICRFPGGVDSPDLMDVVSSEFPDRGMDVDGLDPP 93  
 QY 63 DAPGKTPVTRASTLSVACFDASFFGISPREALRMPARILLVEWGLENAATAPSL 122  
 DB 94 DAPGKTYTRWGAFLDDAAGFDAGFFGIARSEVLAMDPOORLMEVSEWALEYAGIDPLSL 153



QY 6 IERAAEDPIAVGASCLPGVIDLGSFMTLEGSBDTVGVPAERDAAMFDDPDPAR 65  
DB 51 LEAQOEFPIALIMACRFPBGANNPEAFWQLDRDGVAVGEIPDARDIKYDANDPT 110  
QY 66 GKTPTVRASFLSDVACDASFPGISPREALRMDPARLLEVCWEALENNALPALSALT 125  
DB 111 GKMYTRGHPLDYVDEFDQAFPGISPREAQSLDPQOGLLEVSFEALERAGOSPAKIKGS 170  
QY 126 ETGVFISGSEXEALPQATASAEIDAHGGLGMPVSGARISYALGLGPCVAVDTAY 185  
DB 171 QTSIYGLCFDDY--AKHSLSHPTQIDAFSSLGNTSIAAGRIAYVGLGTWQOLDTTC 228  
QY 186 SSSLVAVHLACQSLRSGECSTALAGVSLMSPSTVWLKSTRALADGRCKAFSAEADG 245  
DB 229 SSSLGCHLACQSLRSGESDMALAGVNLILSPBPMGFCKLALADGRCKTLDAADG 288  
QY 246 FPGEGCAVYVYKRLSGARADGRIILAVIRGSAINHDSAGSLTVENGSGOEIVLKALA 305  
DB 289 YGRGEGCGIYVYKRLSRDAIANDPILAVIRGSAVNHDSAGSLTAPNGTAQEAIVIHQALQ 348  
QY 306 DAGCAASVGVVAHGTGTLGDPRIEIQALNAVYGGRDVATPLLGSVKTNLGHPRYAS 365  
DB 349 NAOVEPHQIOYSELHGTVLGDPIEVLAKVGBGRSKNQPLSISVKTNLGHEGAA 408  
QY 366 GITGLKLVLSLOHGQIPAHILHAQALNPRISSWGLRLTVTRARTPMDMNTPRRAGVSR 425  
DB 409 GVALGMKVILALQHQOIPAHVNFQOQPNPYIPWOKLPLTVPTQLTWPQANSQRLAGVSR 468  
QY 426 GMSGTNAHVLEAPATCTPPAP---ERRPBLIVISARTASALDAQAARLDDHLETYP 481  
DB 469 GMSGTNAHVLEAPATCTPPAP---ERRPBLIVISARTASALDAQAARLDDHLETYP 522  
QY 482 SGLGDAVAFSLATRSAMERLVAATS---REGLEAALDAAAOQSTPGAVRSIDSS 537  
DB 523 EDNLDDICTSNGSRSHFPIRLITIPANSYNSVYEQLSNLTNKRKTDS-----QLIPPN 577  
QY 538 RGKLAFLFTQGAQTLGMRGLYDWSAFREARDLQVRLFNQELDRPREVMA--EPAS 595  
DB 578 TLKIAFLFTQGSQYLMARELVDTHPYFRQVDKCELLHPYLDVDRLELFTPERPNS 637  
QY 596 VDAAL-----LDOATFOTPALTFEYALALMRKQVPELYAGHSIGELVAAQVAVF 649  
DB 638 SNELRTNYELSHITITOTOPALFYIEVALTQLMRHMVTPDPFLGHSIGEVAAQVAVF 697  
QY 650 SLDDAVELVABRGIMQALPAGGAMVSIIEADVAABAAPHASVSIIVAVNAPDQVIA 709  
DB 698 TLDPALKILHRLARLWQOLPENGKMLAVSASATEIAEYIOAYSENVGIAINSPEVTVIS 757  
QY 710 GAGQPVHAIAMAARGARTKALHVSHPPLMAPMLAEGRAVESVYRBPISIVLVSN 769  
DB 758 GEQTALIAQLABEILQNGQIKTTPLSVSHAHSPLMETWVEEPQVABESITFHPPTPIISN 817  
QY 770 LSGKACTDEVSSFEYWRHARBEVVRFPADQYKALHAAGAGTFVEVGPSTLGLVAPAMP 829  
DB 818 ITGVIGEETIANEYWMCHILQPOFALGIETLAKONCKIFLEIGVAPRTLSIGKCTCPD 877  
QY 830 ARP-ALLASSRAGDEPATVLEALGMAVAGVSWAGLFPSS--GGRVPLPTPMOSERY 887  
DB 878 AENYMWLPSLRPROSDQOTILSSLASLYMOSIDINMENFTDNOHQHVIILPTTSFOROR 937  
QY 888 MID----- 890  
DB 938 WVDKPEFTPYPPYRNHPLGLQKYLAKSEKIYQNKISPKKDFLOQHQVQNSINPAA 997  
QY 891 ----- 890  
DB 998 GYIEIATTAABOJIPSHSYNLQNIQIOPPTQOETLOILPPLKQOTYQOFELSLCAF 1057  
QY 891 -----TKADDA----- 897  
DB 1058 APLREKETTATNGIITQSSPLPNSPOLPTTCTOKIDPTTYQOTQKLGIEYKGFQALIQ 1117  
QY 898 --RGDRRAPG----- 905

DB 1118 LMRGENQALGEIKLSPEIDTTPYQIHPILILDACLVGTGAIPLDSOPTAVLPIAIEQOFQM 1177  
QY 906 ----- 905  
DB 1178 YGEKIQTHLSYVYKLEREYIADIQIFANNKLIATITGLKIQPVQKLNWOMLYSIAMK 1237  
QY 906 -----AGHD----- 909  
DB 1238 ENPLTYHHLTPKVSNSILNOSLNDVNLSSQLEKSIACILQTLQAGEELSIVTOH 1297  
QY 910 -----EVEBG----- 914  
DB 1298 KKLQYLLQILQNOQRETPLSFTQURQYTPNATSENLKRCCTSHAEVLQGXIDPLQJ 1357  
QY 915 -----GAVRG----- 920  
DB 1358 LFPNGDSULTHLYONSPIAKMMNTLVKQAITTAITDIPPEORPLRIETGAGTGTTAHL 1417  
QY 921 -----DRRSAR-----LDHPPESGREK----- 939  
DB 1418 LPELAQRNIEYFTDVSPLELSKAOQPADYSCVRYQLPDEVEKPTTQGFTEFNSPDIIIA 1477  
QY 940 ----- 939  
DB 1478 ANVLHATEDLRQTIHIKOLLSRGILLILEGVQVYWMIDLIFGLTEGMMRFQDYDLRPH 1537  
QY 940 -----VEAAGDRPF----- 948  
DB 1538 HPLISSEQOKLFTQGFDPSTLLETIGNQAVITTAOTPKLPDNCVILADKQIGEKLATIH 1597  
QY 949 ----- 948  
DB 1598 LOTQGTCHLIFRADSQPTLPTRASHLNLWGLDVHPDELTAELIKQOQLCESLAYS 1657  
QY 949 ----- 948  
DB 1658 LQNNIKYQOLPQMWLVTOGAVNTDQDTSIKGLAATLWGISKVIQLEHBEINCHCLDLP 1717  
QY 949 -----RLEI----- 952  
DB 1718 NHTPIFOLDILKELLTNSPEKEIAMRTNIRYTPRLSONSCLVSAPLREKNSLMORLTI 1777  
QY 953 DEPGVDHLVLRTERBARGLCEVEIAVDAAGLSPNDVOLALGAMPDDLPBGKPNPPLIG 1012  
DB 1778 PQGTLENLFTQVNTSINPGEIERIGATGTFPDIINAL-----DLYRGEBC---MIG 1830  
QY 1013 GECAGRIYAVGEGVNGLVVGOPIYALSAGAFATHVTTSA-----LVLPPOALSAB 1065  
DB 1831 CECVGEIYVAIGTVDKHLQIQCFVIALASGSFSQYVYIKAAAMSPAGIAIPLPQNLNITID 1890  
QY 1066 AAMPAVYITAMVALDIRIARLQGERVLJHAATGCVGLAAVQMAHVGAEVHATCTPEK 1125  
DB 1891 GATTPAFLFAFTLHLKIRGDKVLJHAAAGVGQALIQAKLAGAEIFATA--SPQK 1949  
QY 1126 RAYLESIGVRYVSDSRSDRFVADVRAWTGSEGVVNLNLSGELIDKSFNULRSRGPVE 1185  
DB 1950 WELTRNLGVTIKFNSSTLDPFAEBILQTOGEGVDIVNLNRGDFIAASFVALKPQGRFVE 2009  
QY 1186 LGRKDCYADNOL--GLRPFRLNLSFSLVDL--RGMMLERPARVRLLEBLGLIAAGVTPPP 1244  
DB 2010 IGRIDVWTEBOYQVXP--NANYPLVDLVDLCOQOPEDIIOMLKOLVEEFEOHLOPLA 2066  
QY 1245 IATLPYIARVADAFRSMQAQOHGKLVLTIG----- 1274  
DB 2067 AKVFPNQIVDARFYMOQGHGKVVADWGLNSLYPLKENCYTLITGGIGGLLVAAEY 2126  
QY 1275 ----- 1274  
DB 2127 LVKGAACHLLLGRNAPTAADVASKIOLETTGARIAIQIDVSGTAALSILQIENSAY 2186  
QY 1275 ----- 1274



Db 2187 PLAGIHAAGVLDGALLQNMNERFOKMAKVMGAMNLTHTLPLDFVLFSATATL 2246  
1275 ----- 1274  
Qy 2247 FGSPOGANNVANTFLDTLAHYHTKGLPAMSINMGINSEIGATQATKQMOGRGIEATA 2306  
1275 -----DPEVQIRI-----PTHAGAGPSTGD-----RDJ 1297  
Db 2307 PADGIQIOLHMTQPLTQIGVAVINWQKFPINSPFENFPTN---NPETRQTPPEKTDL 2363  
Qy 1298 LBRLSAARAAARAAALEAFRQVSOVLATPEIKVGAELFTPLGMDSLMAVELRRIRTA 1357  
Db 2364 LELTSLDQYQYELLETLOEQIAYLVGPPHEINPQTFDGLGMDSLTALFEKRLQT 2423  
Qy 1358 SLKTLSTFTLSTPNIALAQLDAL 1385  
Db 2424 DKITLPSTIAPDPYPIQTLAHLHNLNQL 2451

## RESULT 4

H70984  
probable polyketide synthase - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: H70984  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Parkhill, J.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Raftery, M.A.; Rogers, J.; Ruter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G. A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70984  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1602 <COU>  
A:Cross-References: GB:295617; GB:AL123456; NID:q3242249; PID: CAB09098.1; PID:e317262; A:Experimental source: strain H37Rv  
C:Genetic8:  
A:Gene: pks8  
C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
F:55-452/Domain: 3-oxoacyl-[acyl-carrier-protein] S-malonyltransferase homology <AMT1>  
F:1561-839/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

Query Match 30.6%; Score 2207; DB 2; Length 1602;  
Best Local Similarity 35.0%; Pred. No. 2.8e-111;

Matches 555; Conservative 141; Mismatches 460; Indels 428; Gaps 17;

Qy 6 IERRAEDPIAVGASCRLEPGVIDLSGFMTLEGRSDTVGRVPAER-WDAAMFDEPDPA 64  
Db 28 LEAKUSEPAVAVGMCGRPGVDSPETLMBELVAGRDVSDFPADGMDVDGLFDDPDPA 87  
Qy 65 PEKTPTRASFSDVACFPDAPFGISPRBALMDPAHRLLEVCWEALENAATAPSAVVG 124  
Db 88 CCRMYTRRQTFLEHAGDFAGGFGIGPSEBALMDPQORLLEVSWEALERTGIDPTKLRG 147  
Qy 125 TETGVFIGGPSEYEALPOATASAEIDHAGLGSTWPSYGARISVALGRCVAVDTA 184  
Db 148 SATGVPAVYHAGYQGL-----SGELRGYGLTGSTLSVAGRVAVYLGEGPAYSVDTA 202  
Qy 185 YSSSLVAHVLACQSLRSGECSTALAGVSLMSPSTLWLSKTRALARDGRCKAFSAEAD 244  
Db 203 CSSSLVAHLAVQSLRSGECSTALAGVSLMSPSTLWLSKTRALARDGRCKAFSAEAD 262  
Qy 245 GGRGGGCAVVLKRLSGARADGDRILAVIRGSATIHDDAGSSGLTYPNSSQGEIVLKRLL 304  
Db 263 GRAMSEGAVALVERLVDRRLGRHGPVLAIVRGSAVVDGASNGLTAPNPSQQRVTRAL 322  
Qy 305 ADAGCAASVGVYEAHGTGTLGDPTEIOLANAVYGLGRDVATPTLIGSVKTLGHEPVA 364  
Db 323 AARARLAVVDVVEGIGTGTLGDPTEIOLANAVYGLGRDVATPTLIGSVKTLGHEPVA 380  
Qy 365 SGITGLKLVLSLQHQIQAHLHAQALNPRISWGDLRLTVTRARTPWPDMNTPRAGVSS 424

Db 381 AGVAGIYKVMQAMRGVMPKTLHVDVPTPHVDWSVGAASLTQOPRMSVHGSPRRAGVSS 440  
Qy 425 FCMGSGNHAHVLEEARATCTPRAPER-----ALLVYSARTASALDAQAARLRLHLEY 480  
Db 441 FGISGTHAVHLEQAPVESVPEVASPTASAVPWSVARSQEQALAGQORLLFAVAN 500  
Qy 481 PQSCGDVAFSIATRSAMEHRLAVALTSREGRLAALDAAGQOTSFGAVRSIADSSRCK 540  
Db 501 PLDDPIDDGMSLVKTRAMEHRAVVGADRGALLGLAALAGSAGAVGARA-RSVCK 559  
Qy 541 LAFLETGCAQTLGNGRGLYDVWSAFREAFDLCVRLFNOELDRPLREVMMAEPASDAL 600  
Db 560 TVFVFPQGAQVWVGCAQLYAEPLFALAFDAVAEELDRHLPLRNVLW-----EGDEAL 615  
Qy 601 LDQTAFTQPALFTPEYALAAALWRSVGEPELVAGHSIGELVACVAGVSLDEAVFLVA 660  
Db 616 LSTTEPAQALFAIYALATLLQHWGISPDPLIGHSVGEIAAHLAAGVSLTDAGLVAA 675  
Qy 661 RGRMQALPAGGAMYSIEAEADVAAVAAPHAASVIAAVNAPDVVVIAGQOPVYHIAA 720  
Db 676 RGRMAELPAGGAMVVAASVEEVLPLVDGA---NLAAVNPHEVTVSGCEAAVSDIAD 732  
Qy 721 AMAAGARTKALVSHAFHSPIMADPMLAEFGVAVSVSYRPSIVLVNSLSKACTDEVS 780  
Db 733 HPARGRVRHRLAVSHAFFSLMEPLAEFTRIAGISVSKRIPLVSVHTQMGAGAGYG 792  
Qy 781 SGGVAVRAREVVRADQKALHAAGAGFVVGSKSTLGLVPCMPARALLASSRA 840  
Db 793 DGQVVEHARRVRAREGVQQLNAVGAARFVVGQGGTLTALVEOSLPGLGALSVAMMR 852  
Qy 841 GRDEPATTLEALGLMAVAGLVSMAGLPPSGRRAPLPPTYPMQRRRYWID----- 890  
Db 853 EHPVSVVGAATLFTAGQCMDWAVGSPRRRIELPTYARQGRYWLPTISAGSADS 912  
Qy 891 ----- 890  
Db 913 GVLGAARHGLGAVEQDPDSDVVLGTGLSVGEQRLADHVIAGVLLAGAFVELALR 972  
Qy 891 ----- 890  
Db 973 AADQVDCGVEELTVVTLPLPTVGVQQLQVVVGVGEMQRPVSIYSRAESDSCGVLHA 1032  
Qy 891 ----- 890  
Db 1033 RGVLAQKAVAPAADSWMPPLGAAPVDVDAVQRAELGYEYGRAPQGLTAMRRESELP 1092  
Qy 891 ----- 890  
Db 1093 ADVAVPDVDVTLTSGFGIHLPLVLAALHAMGVGQAAATMLPFSMOGVSLHAAGASRVRA 1152  
Qy 891 -----TKDDAARG----- 899  
Db 1153 RIAPAGDGTVELDQGLPVLVQVALVMSVSSQLSAAVAADADAGRLLEVAMLPV 1212  
Qy 900 -----DRAP-----GAGHD 909  
Db 1213 ELAHDIDSLDVVLESTQDQGVGYATHRVVLALQSNLAQBARGLVVLTOGSGVD 1272  
Qy 910 EVEEGAVRGDRSARLDHPPESGRREKVAAD----- 945  
Db 1273 ATNLGAAGVAGLVRSQAQEHF-----GRVWLVDSDSMQDGVIGCGEQLMIRNGTAYAA 1328  
Qy 946 -----RP-----FLEIDEPVLDHLVLRVTERRAPGLGEVEIADAGLSRN 988  
Db 1329 RLQAQRPPPIQLPDTNSGMRVAVAGAGALBDLTLASCPAKLABGQVRIETRALGVNFR 1388  
Qy 989 DVQLAGWPPDLPEKPNPPLTIGGECAGRIYAVEGVNGVLVGOPLVLSAGAPATHT 1048  
Db 1389 DVLVALGTPGAAL-----LGAEGAGVTEVEPGTGLAVGDPWGL-LGVAGSEAV 1439  
Qy 1049 TSAALVPRPOLSAIEAAMPVAVLTAVYALDRIARLOPGERVLIHAATGSGVLAAYVW 1108

Db 1440 VDARLVKLPJNRWPLTDAAQPVVFLTAAYVALRVAQVOPGESVLYHAAGVGMAVQL 1499

Qy 1109 AGVGAEVHATACATPEKRAVLESIGV--RYVDSRSRFRVADYRAVMGSGVUVYVNSLS 1166

Db 1500 ARLMGLEVFAYTA-SRGMWDTLHTMGCDNTHVADSRTLAETFMLTTEGGVADVNLSTA 1558

Qy 1167 GELIDKSFNLRSHGRFVELGKRD 1190

Db 1559 GEFTDASLRLLPRGGRITEMGKTE 1582

RESULT 5

A12140

polyketide synthase [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C:Accession: A12140

R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S

DNA Ref. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: A12140

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2518 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA074379.1; PID:gl7131773; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr2680

Query Match 29.6%; Score 2136; DB 2; Length 2518;  
Best Local Similarity 24.7%; Pred. No. 3.5e-107;  
Matches 622; Conservative 246; Mismatches 515; Indels 1132; Gaps 38;

Qy 8 RAAEDPIAIVGASCRPLGVIDLSGFWTLEGSRTVGRVPAERMDAAAFDDPDPAKG 67

Db 24 KAQTEPIAIVGMCGRFPAGADTPEKLMQLLRNGDDGITEIPSDRWSIETFPNNPATRGK 83

Qy 68 TPVTRASFSLDVACFDASFGISPRELRNDPAHRLLEVCWEALNAAIAPSLVETET 127

Db 84 MSTRYGGFIGHLKEFADPFGLAPKEAISLPQORLLEVTWALENTGILPBOITSSQT 143

Qy 128 GVFIGPSEYBAALPATASAEIDAHGGLTSPVAGRIAYALGKGFVAVDTYSS 187

Db 144 GVFIGSSNDYTHLNRPVY-DIDAYLATGNSHSTAGRISTYLGFTGSLAVDTACSS 202

Qy 188 SLVAVHLACQSLRSGECSTLAGVSLMLSPSTLWMLSKTRALARDRCRAFAEADFG 247

Db 203 SLVAVHLACQSLRSGECQQLAVGVGNRI FSPSEFTINFSQARMLAADRCRKTPTDKAGFV 262

Qy 248 RGGCAVAVLKRSLGAAADGRIILAVTSGAINHDAASSGILTVNGSSQEIYLRALADA 307

Db 263 RGGCGVYIIQKLSDAQANPILAVTRGSAVNDGRSSGILTVNGSAQQAQVIRQALKNA 322

Qy 308 GCAASSGVYBAHGTGTLGDPTEIOALNAVYGGRVATPPLIGSVKTMGHEVYASGI 367

Db 323 TVEPSDITVEAHGTGALGDPLEIGALGAVGESHQ-TSPLVLGSLKTNIGHLEAAAGI 381

Qy 368 TGLKLVLSLQHGQIPALHLAQAALNPRIISWGDRLTYTRATTPWMDNTPRAGVSSFGM 427

Db 382 AGLMKVVLALKHGIPIPHLHQPNPHIPMEQPIQIPPTCTP---WHGKTLAVSSFGF 438

Qy 428 SGTAHVAVLEBAPATCTPP-APERPABLVLAKRTSALDAQAARLRDLHETPSCQCG 486

Db 439 SGTAHVAVILAASEIONKENTYIKRPLOIITLSAKTETALKQALQANYITYLKEHQDINLA 498

Qy 487 DVAFSLATTSRA-----MEHRLAVAA-----TSRGLRAALAAAGOGSPGAVRSIA 534

Db 499 DICYSANTGSSRRGAASLTOPEHRLSTIANSITTAOEKLTICFINNEATNMLPTNKVMS-- 556

Qy 535 DSSRGKLAFLFTGGAQTLAGKRGILYDVWSAFREAFDLCVRLF--NOELDRPLREVMVAE 592

Db 557 -PSRPKIAELFTGQSGQYQONMGWELTYQTEPFKEKTDKCCCELLANSGID--LHSVLPQ 613

Qy 593 PASVDAALDQFAFPTOPALFTEPEYALALMRSGVPELVAGHSIGELVAACVAGFSLTE 652

Db 614 DNENHKLISQYLYTPPALFLEIYALCOLMISGLIPDPRMGHSGVEYVAACLAGFSLTE 673

Qy 653 DAVFLVAARGRLMQUALPAGGAMVSIAPRADVAAVAAPAAVSIAVAAPDQVVIAGAG 712

Db 674 DGLKLIATBARLMQQLPQ-GKRVAAVASSQQLNPLFPDQGVSIANAIPNNTVISGEF 732

Qy 713 QPVHAIAAAMAAAGATKALHSHAFSPALMELAFGRVAVESVYRPSVLYVNSLG 772

Db 733 AAIEKIIAVLSQNIQATPLSVSHAFSPMETMLGEFEKIAATINPHRPKYPIISNVIG 792

Qy 773 KACTDVSSPGYVVRARAEVVRADGVKALHAAGATFVEVGEKSTLLGLVRPQMDAR 832

Db 793 SLINSLIAPPEYECRIKQPVQPLAGVETLIQONCSI FLEVGAKPILLOMANSIISQADK 852

Qy 833 AL-LASSRAGRDEPATVLEALGGLMAVGLVSWAGLFPSS-GGRRVPLPTYPMQREYVWD 890

Db 853 YMLPRLQPKQDQWQVMLSSIAAIYGRGVKIDHMKFEQVYHHQRDLDPVYFQROQFWID 912

Qy 891 TK----- 892

Db 913 IKPVNKVYSSSPNNIHPILGQQLNLAKSSDIYEQOULTNNPBDYLQDHKVFNQIILPGA 972

Qy 893 ----- 892

Db 973 AYLEMSLAAKTIFKNNSSIVETVSGFQECILTPDPNKAIQPIKNNHEFIYVTSNMD 1032

Qy 893 ----- 892

Db 1033 WITHATGKIKPHNNIPQPOINLNGELQNTFTKIITIDTFYQNLQVGIEXGETFOATIH 1092

Qy 893 --ADDAAGDRAP----- 904

Db 1093 WYLDNOLAEIHLPSTCNCRDNYQFPHIILDACTQITAAIFYSQPTPSNSSODIENPSN 1152

Qy 905 ----- 904

Db 1153 LSPKRRBALNPSPVLGKVGRLGLTNTNOVYLPIDVDKLTILYPIGETWMSLVKLRTOH 1212

Qy 905 ----- 904

Db 1213 KTAFAVADIQILSPSGOVIALVEGLQKKIQSPNIKDMQNMLEYEIMRSQPLSITSDIS 1272

Qy 905 ----- 904

Db 1273 LTSPEISQELAKFTELLTTEBEIKTYAQLLPQLEALSUYIVQALQNLVDLSPEIAPQH 1332

Qy 905 ----- 904

Db 1333 QRLYHLISLITNOREYKRPQLSPQHPIAEELTIERCGTNLGVLOKCKNPIDLLFP 1392

Qy 905 ----- 907

Db 1393 NGDLTTLTQYXNSPGAKVNTLVQAINSALKDLPRGEKVRILEGAGTGTAAYVLPQ 1452

Qy 908 --HDEVE----- 912

Db 1453 LVHQSVSEYFTDISPLAKARQOFSEYEFVSYOTLINIEQPLTNODITPHSFDIVIAAV 1512

Qy 913 ----- 919

Db 1513 LHATENLNTITNVKSLNNGGILLVBEGITPBIWIDLFGLTGEMWRFPQDRLRPHHL 1572

Qy 920 ----- 919

Db 1573 ISTTAHSLKTHDFTNINIVITPDSILPEALAQSVIAQNTPPQSPYKREGECLIT 1632

Qy 920 ----- 919

Db 1633 DLPANGAIALNQQKLPSPFLKPSDNIAPFPHOSIKHIIYIACQDNITECNNILHLVQTL 1692  
QY 920 -----GDRRSALDHP----- 930  
Db 1693 IKTQHYPIINMLVTOGAISPHTITGLNOSWGMKVILHEPBLNCRVDLDPQOELFT 1752  
QY 931 -----PPESGR-----REKVEAADRPFRLEIDEPVLDHLVLRV 965  
Db 1753 QVENLVTEITHPGEAEQITLHKSERKVARLCTQILPHEPYRLTIARGVYASLKMOS 1812  
QY 966 TERRAPGGEVEIANDAGLSFNDVQLAGMPPDDLPKGRPNPRLILGGCAGRIYAVGEG 1025  
Db 1813 SRRRQPGQEEVERIQVATGLNFIDVLDLGL-----LPFRN-----WFGVCAEEVVAIGGG 1865  
QY 1026 VNGLVGQFVIALSAGAFATHTTSSALVLPRLPQALSAIEAAMPAYVITAMVALDRLAR 1085  
Db 1866 VTHLVGDAVVALADSFQYVTTNANYIKKPDLSFTAAATIPANFTIAYALREYAK 1925  
QY 1086 LQGERVLIHAATGVGVAIVQMAQHVGAENHATAGTPERKAYLSLGVRYVSDRSDFR 1145  
Db 1926 IQPGRKIILHAATGGTGMALQIAQOAGLEVPAIASV-GKMETLRALGVQHIENSRITNF 1984  
QY 1146 VADVAMNGEGEVADVNLNSLSEGLDKSFNLRSRFPVLEKRDY-ADNQLGRPFGR 1204  
Db 1985 AEEHETIQEGEVADVFNLSGSEFIPASLSVLKPGQFIEIGKGMVAQKVAQVFP--- 2041  
QY 1205 NLSFSLVDLRGMLEBRPARVRLLEELGLIAAGVF-----TPPIATLPIARVADA 1256  
Db 2042 DVVYHLVDLMSVAQOQFOTITQTLHCLMAEFESGELRMSPTRRSHPTITPATKIYEA 2101  
QY 1257 FRSMAQOHLGLVLT----- 1272  
Db 2102 LQMQOARHIGKIVITHTPTESLQPDATVYLTGKMGGLGRVAVMLVEKAGHVLVLRN 2161  
QY 1273 ----- 1272  
Db 2162 QPDITTAQOQIALEATGAKIITTOADVSQKQDLAAVLTDIQONHPLRGVTHAAGVLDDG 2221  
QY 1273 -----LGDPB----- 1277  
Db 2222 ILQQLTPERSHRVMPKVTGAMNLIQTLQDIPLDYFIMSSAASLLGSGQANHYAANTF 2281  
QY 1278 ---VQIR---IP-----THAGAPSTGDR-DLID--- 1299  
Db 2282 LDALQVYRRHGLPALSTINMGVMSDIGAAKQVSNQMSRGIGETTPQGDILIEFLIT 2341  
QY 1300 -----RLAS-----AAPAA-----RAAA 1312  
Db 2342 QSTGVGVPIINWSELKQKLSSTFPADFEYESTPSTQSEYQPSOLIOTICIGIKERISY 2401  
QY 1313 LEAFRTQVSOVL-----RTPEIKVGAELFTRLGMDLSMAVELNRIEASLKLSTTF 1367  
Db 2402 LKRIQTEVSQVLGSSSSQLPWNQUG---FPDMGMSLMMVELSRLETSLKQITPSTV 2457  
QY 1368 LSTSPNIALLAQNLDAATALSLER--VAAENLRAGVQNDPVSSGADDMETII 1420  
Db 2458 LPEHRSI-----QALAEVIATQLLPQETVISAINEQPDISSVSSKQEIETSIITA 2508

RESULT 6  
AB2012  
hypothetical protein all1648 (imported) - Nostoc sp. (strain PCC 7120)  
C.Species: Nostoc sp. PCC 7120  
A.Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C.Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C.Accession: AB2012  
R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irituguhi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A.Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A.Reference number: AB1807; MUID:21595285; PMID:11759840  
A.Accession: AB2012  
A.Status: Preliminary

A.Molecule type: DNA  
A.Residues: 1-1587 <KOR>  
A.Cross-References: GB:BA000019; PIDN:BA878014.1; PID:g17135468; GSPDB:GN00179  
A.Experimental source: strain PCC 7120  
C.Genetics:  
A.Gene: all1648

Query Match 29.4%; Score 2120; DB 2; Length 1587;  
Best Local Similarity 44.7%; Pred. No. 1,4e-106;  
Matches 450; Conservative 168; Mismatches 337; Indels 52; Gaps 14;

QY 6 IERRAEDPIAIVGASCRLLPGGVIDISGFWTLLBSRDTGVRVPAERMDAAWFDPPAP 65  
Db 26 LERTQNEPIAIVGMCRRPPGDANNPEKEMELLRQCKDITTPPQRMIDAYDDDDPVP 85  
QY 66 KCTPTTRASFLSDVACRPAASFPGISPRALMDPANHLLRVCWALENNAIAPBALVGT 125  
Db 86 NKMVARVGGFTINNVQFDPQFEGITPRBAIALDPQRLLEVSMALENAGIAPQKLTGT 145  
QY 126 ETGVFIGIGPSEYEALPQATASAEIDAHGIGTWPVGAGRIISYALGIRGPCVAVDTAY 185  
Db 146 QTVGVFVIGIDY--AKQIKHHPIDAYTSGNACFCAAGRLSYLLGLQPSLAIDTAC 203  
QY 186 SSSIVAVHACQSLRSGECSTALAGVSLMLSPSTLVMLSKTRALARDGRCAFSAAEDG 245  
Db 204 STSLVTHLACQSLRNGECNLALAGVSLMLSPVTLVLSKTRALSPDGRCKTFDRDANG 263  
QY 246 FGRGGCAVAVVULKRLSGARACDRILAVIRGSAINHDCASSGLTYPNSSQDEIVLKRALA 305  
Db 264 YVRGCGCMVVULKRSSAVADGDHLAVIRGSANVQDASSGLTYPNQTAQDAVIROLA 323  
QY 306 DAGCASSVGYEAHGTGTGDPDIEIOALNAVGVYGLGRDVATPPLIGSVKTNLGHPEYAS 365  
Db 324 NAKTPPAQISTLEAHGTALADPLEVATIDGVRKGSPPNHPLILGSKTNIGHLEINA 383  
QY 366 GITGLKTVLSLQHQIPAHVLAQALNPRISMGDLRLTVTRARTPWPMNTPRRAGVSF 425  
Db 384 GMAHLKYILSLQHQEIPPHINPQELNPDLAASAKSLKIPISYIMQTEGRMAGISSF 443  
QY 426 GMSCTNAHVLEEARAATCTPPAPRPAELVLSKRTSALDQAQALRDHLETPSQCL 485  
Db 444 GLSGTNNAHIIIEBPQLVTTPAEVDRPLHVLMLSKSSAALHTTLATDWEHLRHNPTNF 503  
QY 466 GDVAFSLATTSAMEHRLAVALATSRREGRLALDAAQOCTPGAVRSIADSR--GKLAFL 544  
Db 504 ADLAFSANTGSGFPHRLAIIAQSTAQARKVLAENQQLPBLNVSQVEVKGROCKIAFL 563  
QY 545 FTGQGAQTLGNGRGLYDVMSAFREAFDLCVRLFNQELDRPLREVWMAEPASVADALDQT 604  
Db 564 FTGQGSQYVVGMRQLYETQPTFRQALDECDRLDQYLKESLSLVLPQTPTAN-PLINQT 622  
QY 605 APTQALPTFEYALALMRSVNGVEBELVAGSIGELVACVAGVSLDANFLVAARGRL 664  
Db 623 AYTOTALPAIETVALCKLMQSWGIOPOGVLGSVSEYVAACTAGVYTLQEGIELIAQORQL 682  
QY 665 MOALPAGGAMVSIAPREADVAAVAAPHAASVIAVANAPDOVYTAGAGOPVAAIAAAMA 724  
Db 683 MOALPQITMAAVPAPVETVARAIAPYANETIATINSPEANVYSGVAAIALVADULTA 742  
QY 725 RGARTKALVSHAPSPMLAEMLEAFGRVAESVYRPSIVLVNSLSKACTDEVSSPGY 784  
Db 743 QGIDVRPLQVSHAFHSFPMWEMLEPFKQVAAKINQOTGIDMISVTAETHTSIDAE-Y 801  
QY 785 WVRHAREVVRADGVKALHAAGAGTFVVGKSTLILGLVPCMPDARALLASSAGADE 844  
Db 802 WCOQIRDCVQAPAMMETLAQOGYDVLEIGHVPLVTRGSKTSLSPQITLWPSLHRENN 861  
QY 845 PATVLEALGMAVAGVLSMAGLFPSSG-GRVPLPTYMORERYWIDTKADDAARGDRA 903  
Db 862 WQSLQSVATLSTVNGVRLDMSGFEODYIRRLITPTTPFOQRVWL-----AAESTQ 914  
QY 904 PGAGHDEVEEGAVGGDRRSARLDHPPESGGRREKVEAAGDRPFRLEIDEPVLDHLV 963

```
Db 915 P-----EVIPVAAL-----SPET---STIVA-----TETLESQILS-LVA 946
Qy 964 RYTERARPGLGEVEIAVDA---AGLSFNDV---QLALGMVPDDLPGK 1004
Db 947 KIT-----GMMPOQLSLDALTLEGGLGDSIMTMQTMNGILIKTIPQ 988

RESULT 7
S23070
erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea
N/Alternate names: 6-deoxyerythronolide B synthase II
C/Species: Saccharopolyspora erythraea
C/Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text_change 21-Jul-2000
C/Accession: S23070; S23011; S23205
R/Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. 204, 39-49, 1992
A/Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of
A/Reference number: S23070; MUID:92155230; PMID:1740151
A/Molecule type: DNA
A/Residues: 13573 <BEV1>
A/Cross-references: EMBL:X62569
R/Bevitt, D.J.
submitted to the EMBL Data Library, September 1991
A/Reference number: S22011
A/Molecule type: DNA
A/Residues: 1-184, 'T', 186-301, 'S', 303-521, 523-658, 'A', 660-993, 1001-1212, 'H', 1214-1392, 13
3479, 'DH', 3480-3572 <BEV2>
A/Cross-references: EMBL:X62569; NID:946977; PIDN:CAA44448.1; PID:G581651
R/Gaffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.
FEBS Lett. 304, 225-228, 1992
A/Title: Identification of DEBS 1, DEBS 2, and DEBS 3, the multienzyme polypeptides of th
A/Reference number: S23103; MUID:92316235; PMID:1618327
A/Accession: S23205
A/Molecule type: protein
A/Residues: 2-12, 'XXX' <CAF>
C/Experimental source: strain CA340
C/Genetics:
A/Gene: eryA
A/Start codon: GTG
C/Function:
A/Description: catalyzes the construction of a polyketide chain, which is then cyclised
A/Pathway: erythromycin biosynthesis
C/Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car
ort-chain alcohol dehydrogenase homology
C/Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi
F/52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F/561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F/1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F/1404-1475/Domain: acyl carrier protein homology <ACP1>
F/1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F/2023-2303/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F/2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>
F/3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>
F/3420-3493/Domain: acyl carrier protein homology <ACP2>
F/1139/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 29.0%; Score 2093.5; DB 2; Length 3573;
Best Local Similarity 32.3%; Pred. No. 1,1e-104;
Matches 607; Conservative 178; Mismatches 532; Indels 561; Gaps 45;

Qy 2 ADPRIEPAE-DPAIVGASGRPLGVIYDLSGFMTLL-EGSRDVGKVPAR-WDAAMP 58
Db 1487 AVRVAADSESEPIAIVGICRPGGIGSPQQLRWVLAEGANLTG-PPADRGMDIGRLV 1545

Qy 59 DDDADGKTPVTASPLSDVACFDASFPGISPREALRMDPAHRLLEVCMEALENAIA 118
Db 1546 HPDDNDGTSTVVDKGLTDAADFPPGFITPREALAMDQRLMETANEAEERAGCID 1605

Qy 119 PSALVGTETGVFIGPSEVEEALPQATASAEIDAHGGLTGPSPVGARISYALGRPC 178
Db 1606 PDALRGTDIGVEVMNGMSYQML--AGEAEKVDVYOGIGNSASVLSGRIAVTFMEGPA 1663
Qy 179 VAVDTAVSSSLVAVHLAAGSLSGCSTALAGVSLMSPTLVMLSKTRALABDRCKA 238
Db 1664 LVVDTRACSSSLVGIHLAMQALRGCSTLACGVTVMSPPTVFVDFSTRGLASDRCKA 1723
Qy 239 FSAADGFGRCGCADVVLKRLSGARADRIILAVIRGSAIHHDSAGSLTPVNGSSQEI 298
Db 1724 FSAADGFGALSGEVALVLEPISSRAANHQVLAVIRGSAVVDGASNGLAAPNGSQER 1783
Qy 299 VLKRALADGCAASSGVYEAHGTGTTLCDDPIEIQALNAVYGLGRDVATPLIGSKYTL 358
Db 1784 VIKQALAAAGVPAADVVEAHTGTETLDDPIEAGLITVYGODRD--RPLIGSKYKTI 1841
Qy 359 GHPEVASGITGLKVLSTHQGIPAHILHAQNLNPIISGDLRLVTRARTPMPDMNTER 418
Db 1842 GHQAAAGAGVYKVLARHQLPRSLHADELSPIIDESGAVEYREVPMPACERR 1901
Qy 419 RAGVSSFGSGTNAHVLEAP-----AATCTPPAPERAEILLVARSASALDAQARL 473
Db 1902 RAGVSSFGVSGTNAHVLEAPAEQAEARTERGPIP-----FVLSGSEAVVAQAARAL 1955
Qy 474 RDHLETYPQCCGDVAFSLATTRSAMEHRLVAATSRBGLRALDAAAGCOTSPGAVRSI 533
Db 1956 AEHLRDTPELGLTDAAMTLAGRARDVAAVLDGDRAGVCAELDALAGRPSADVAVP 2015
Qy 534 ADSRRGKLAFLFTGOGAQTLLGMRGLYDVWSAPREAFDLCVRLFNQELDRPREVMAMP 593
Db 2016 TSPAPR-KPVLVFPGQCAQVGNARLDLSESEVFAEBSMCAERLSHTHTKLLDVVRGCG 2074
Qy 594 ASYDALDLDTAFTQPALTFEYALAAALRSWGVBEELVAHSGISGLVAACVAGVSLD 653
Db 2075 GPDHPRVD---VLQVLFVSIWLSLELRAHGVTPAAVGVSGQEIAAAHVAGALSLBA 2131
Qy 654 ANFLVAARGLMQALPAGAMVSTIEAPEADVAANAAPHAASVIAVAPDQVITAGAO 713
Db 2132 AAKVVALRSQVRELDDQGMVSVGASRDELTVALRMDGRAVAAVAVNPGTGVAGPTA 2191
Qy 714 PVHAIAAAMAAGARTKALHVSHPSPMAAPMLBAFGRVAESVYRPSIVLVSLSGK 773
Db 2192 ELDEFFAEAEAMKRRRIAVRYASHSEVARIEDLAELGITTVAVRSGVPLHSTVTGE 2251
Qy 774 ACTDESSSPGVWRRHAREVRFADGVKALHAAGAGTFVEVGPKSTLL-----GLV 823
Db 2252 VIDTSMADASWYWRNLRRPVLFEQAVRGLVEQCFDFVEVSPHPVLMAVEETAEHAGAB 2311
Qy 824 PACMPARPALLAASGADDEPATVLEAL--GGLMAVGLVSAAGLP--SGRRARPPLPT 879
Db 2312 VTCVP-----TLRREOSGPHEFLRLRAHVHVG-----ADLRPAVAGGRPAELPT 2358
Qy 880 YPMQREBYM----- 888
Db 2359 YPEHOFEMRPHRPADVSALGVGAEHPLLAADVPGHGAVFGRGLSTDEQPLAEH 2418
Qy 889 ----- 888
Db 2419 VVGRTLVESVLDLAAAGEDVGLPVLBELVLQRPVLVLAGAGALLRMSVGAPDESGR 2478
Qy 889 -IDTKADD-----AARGR----- 901
Db 2479 TIDHAEDVDLADAOQSHATGTLAQGVAAAGPRTEQWPPEDAVRIPLDDHYGLAEQG 2538
Qy 902 -----RAGAGHDEV-----BEG-----GAV----- 917
Db 2539 YEYGPSEFQALRAWRKDDSYAEVSLAABEGYAFHPVLLDAVAQTLSIGALGEPGGKL 2598
Qy 918 -----RGGDRSARL-----DHP 931
Db 2599 PFAMNTVTLASGATSVRVATPAGADAMALRTDPAGHLVATVDSLVRSTGEKMQEPE 2658
Qy 932 PESGRRE-----KVEAAGRP-- 947
Db 2659 PRGGEGBLHALDWRLAEPSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPRA 2718
```

QY 948 ----- 947  
Db 2719 AARHGVLMAALVRRLLEQELPGATLVATSGAVTVSDDDSVPEEGAAAMCVIRCAQA 2778  
QY 948 -----FRLEID-EPGVL-----DHLVRY-----TERRAPG- 972  
Db 2779 ESPDRFVLLDTPAEPMPLPVPDPNQLATRGDDVFVPRLSPLAPSLTLEAGTQRDVPGD 2838  
QY 973 -----LGEVELAVDAAGLSFMDVQALGMDVDDLPKKNPFLILG 1012  
Db 2839 GAIDVAEPAPDVEQPLRRAGSVRVDPATGVNFRDVLALGMYQOKAD-----MG 2890  
QY 1013 GECAGRIVAVGEGVNLVVGQPIYALASAFATHVTTSALVLPFPQALSAIEAAMPYA 1072  
Db 2891 TEAAGVTVAVGPVDVAF-PGDRVLGFGQAFAPATDHRLLARVGDGSDADAAVPTA 2949  
QY 1073 YLTAMVALDRIALRLOGERVLIHAATGVGLAAVQAQHVGAENVHATAGTPEKRAYLES 1132  
Db 2950 YTTAHVALHDLGLRAGQGVLIHAAAGVGMVAVALARABAEVLATAG-PAKHGTLRAL 3008  
QY 1133 GV--RVSDRSRFRPADVRAWMTGEGVUVVNSLSEGLIDKSFNLRSHGRFVLEGRD 1190  
Db 3009 GUDDERIASRRETFARKFRERFGVGVVNSLTGELLDESADLLAEDGVFVEMGKTID 3068  
QY 1191 CYADNQLGLRPFLRLNLSFSLVDLGRMMLEBPA-----RVRLLEELGLIAAGVF 1240  
Db 3069 -----LRDAG-----DRG-----RYAPFDLGEAGDDRLGELLEVGLLGAAGEL 3108  
QY 1241 TPEPIATLPIARVADAFRSMAQOHLGKLVLTG--DPEVOIRIPTHAGAGSTGRDL 1297  
Db 3109 DRLPVSAMELGSAFALQHMRSRGHVGMVLTQPAVDPDGVLI-----TGCTGT 3159  
QY 1298 LBRLSAARAARAAALAEFLRTOVSQVLTREPETKVGAELFTPLGMDSLMAVELRRIRA 1357  
Db 3160 LGRLL-----LARKLVTEHGR-----HLVSRGADACGSBELRAEIB- 3197  
QY 1358 SIKLKTSTFTLSTSPYIALAQNLDALATL-----SLERVAEN-LR 1400  
Db 3198 DLGASAEIACDTADADALSA--LUDGLRPLRTGVVHAAGVADGLVTSIDERAIVEQVLR 3255  
QY 1401 AGVQNDVFSSGADQWEI 1418  
Db 3256 AKV-----DAANWL 3264  
RESULT 8  
T17467  
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei  
C:Species: Amycolatopsis mediterranei  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 17-Nov-2000  
C:Accession: T17467  
R:Schupp, T.  
submitted to the EMBL Data Library, December 1997  
A:Reference number: Z18802  
A:Accession: T17467  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-3413 <SCH>  
A:Cross-references: EMBL:AJ22012; NID:e1227119; PID:e1227124; PIDN:CAA11039.1  
A:Experimental source: strain LBG A3136  
C:Superfamily: acyl carrier protein homology  
C:Keywords: carrier protein  
F1:1608-1679/Domain: acyl carrier protein homology <ACPI>  
F1:3334-3405/Domain: acyl carrier protein homology <ACP2>  
Query Match 28.6%; Score 2063.5; DB 2; Length 3413;  
Best Local Similarity 35.1%; Pred. No. 4.3e-103;  
Matches 601; Conservative 195; Mismatches 547; Indels 369; Gaps 56;  
QY 8 RAADPDIAIVGASCRIPGVYIDLSGFWTLEGRSDTVGRVPAER-WDAAMFPDPDADG 66  
Db 27 RHADDELAIVGACRPPGVSSPEDLMQIVAGGVADLSDFPDDRGWELDGLFDPDPDHPG 86

QY 67 KTVTRASFSLDVACFDASFFGISPREALRMDPAHRLLEWCWEALENNAIAPSLWGT 126  
Db 87 TSYTQGGFLRAGLFDAGLFCISPREALVMDPQORVLESTWEALDEDGCVPLSLKSGD 146  
QY 127 TGVFTIGIPSEYEALPQATASAEIDAHGGLGTMPSVGAGRISVALGLRGCVAVDTAYS 186  
Db 147 VGVFSGVFTQGYCA-----GAIRPDLAEAFGIGAASSVSGRVSYVFGLEGPAVTTIDTACS 202  
QY 187 SSVLVNHLACGLREGECESTALAGVLSLSTLWLSKTPALARDGCKPFSAEADDF 246  
Db 203 SSVLVNHLAQAQLRAGECSMALAGATVPTGTVAASRQVRLAADGSKAFSSADDT 262  
QY 247 GREGCAVVLRLTSGABADGRILAVIRGSAIINHDSASSGLTPNGSSQEIYLRALAD 306  
Db 263 GMAEGAGVLVLERLSVADGERHRIILAVIRGSVANDGASNGITANGPQGVIRKALLAG 322  
QY 307 ACCAASSGVYBAHGTGTTGDPRIEIQALNAVYGLRDVATPLLIGSVKTNLGHPEYASG 366  
Db 323 AGLVASDVQVVEAHGTGTLGDPRIEIQALNAVYGLRDVATPLLIGSVKTNLGHPEYASG 380  
QY 367 ITGLLKVYLSLQHQIIPALHQAOLNPRISWGDLLTYTRATPPMDMTPRRAGVSSRG 426  
Db 381 VAGVIAKMQVALRHGAMPPTLHAABPTPEVDMSAGAVELTREPRPAGRPRAAGSARF 440  
QY 427 MGTNAHVVLLEAPAPATCTPPAPERPAEL-LVLSARTASALDAQAARLDHLETYPSOCL 485  
Db 441 IGTNAHLILEAPAPADVAABEPKGPVLYVSGSPSLAQAQRLAEVLAS----- 494  
QY 486 GDV-----AFSLATRSAMEHRLAVATRSREGRLAALDAQAQOTSPGAIVSIADSSRG 540  
Db 495 GGVSAARLASGLLSGRALLGDRVAVVAGTDEDAVAGLRALARGDRAFGVLTGSA--KHKG 552  
QY 541 LAFLTGCAQOTLGMGRGLYDVMSAFREAFD-LCYRLRNOELD-----RPLAEVMA 591  
Db 553 VYVVFPGQSQRLGMGRGLYDYRYPVFATAFDAC-----EQLDVCLAGRAGRVDVVLG 607  
QY 592 E-PASVDAALDQTAFTOPALFTFEYALAAALRSWGEPELVAHSGISGLVAAVAGVS 650  
Db 608 EVPA--ETGLNQYFTTQAGLFAVESALFRLAESGVRPVDVLGHSIGITAAVYAGVS 665  
QY 651 LEDAVFLVAARGRLMQALPAGGAMVSIAPADVAVAAPHAASVIAAVNAPDOVVIAG 710  
Db 666 LPDAARIYAARGRLMQALPAGGAMVVAASAEVALELGD--GVELAANVGPSAVVLSG 722  
QY 711 AGQPVHAIATAAAGAARTKALHVSHTSPMAEMLEAFGVAASVSIRRSIYLVSL 770  
Db 723 DADAVVAARMRERGHKTQKLVSHAFISARMAEMLEFAELAGVTRBEPIPVVSNV 782  
QY 771 SGK-ACTDEVSFGVWRHAREVVRPADGVKALHHAAGTFVEVGPDKSTLLGLVPAQMPD 829  
Db 783 TRFAPBGEELTPGVAEHVRPVRFAEGVAATAATSGSLFELBPGALTYLVE---ET 839  
QY 830 ARPALLASRGRDRDEPATVLEALGILMAVGLVSWAGLPPSGGRVRLPTYPWQRRRYWI 889  
Db 840 AEVTCVAALRDRDRPVTAITAVAEIPLRGVAVDWPALBPVPTGFDLPKVAFDQHWYL 899  
QY 890 D--TKADDA-----RGDRRARGA-----GHDEVE---GC----- 915  
Db 900 QPAADATPAASIGVAAAHPIPLGAVVRLPQSDGLVTSRLSLKSPHMLADHIGGVVLA 959  
QY 916 -----AVRGDRRSARL-----DH-----PPESGR--EKV 940  
Db 960 GIGVELAVRAGDEAGCPVLELVTEAPLVVDHGGVRIQVNVGAPGEGISAVEVYSIR 1019  
QY 941 EAAG-----DRPF-----RLEIDE--PGVLDH-----L 961  
Db 1020 EDAGAEVNAARHATGFLAATPSQHKPFDFAMPPEVERVDEVDYDGLVDGRYAGPSFR 1079  
QY 962 VLRATERAPGL-GEVELA-----VDAAGLSFN-----DVQALGM---VPDDLPGKPNP- 1008  
Db 1080 GLRAVWRGRDEVFAEVALAEDDRADARFGILHGLLDALAHGAGGATTEBGRBPVLPF 1139



```

Qy 930 ----- 929
Db 2709 DFTLDALHLEWTAMDGAAQALPGAVVVGDDADGLAAALRAGSTEVLSFPDLTDVEAND 2768
Qy 930 ----PPES----- 934
Db 2769 RGETAPATATVLAACPAGPDGPEHVEALHSLALQWMLADERFTDGLVTRDAVA 2828
Qy 935 -----GREKEA-----ADRP----- 947
Db 2829 RSGDGLRSTGQAAVWGLGRSAQTESGFEVLLDLAGEARTAGDTAGDTTGDATVGGT 2888
Qy 948 ----- 947
Db 2889 SGDAALGSLATRALSGEPQALRDGALLVPRARAAAPAAADGLAAADGLAALPLPAAP 2948
Qy 948 --FLEIDEPVLDHLVLRV--TERRAP--GLGEVEIAVDAAGLSFNDVOLALGMVPPD 1000
Db 2949 ALMRLEPGTDSLESLETAAPGDAETLAPPLGPGQYRIRATGILNFRDVLIALGMYPD- 3007
Qy 1001 LGGKPRPILLGEGCGRIVAVGEGVNGLVGQPVIALSAGAPATHVTTSALVLPBQA 1060
Db 3008 -----PALMGTEGAVTATGPGVTHLAPGDRVWGLLSGAVAPVADARTVAMRPG 3060
Qy 1061 LSAIEAAMPVAVLTAMVALDRIARLQPERVLIHAATGVGLAAVQMAQVCAEVHATA 1120
Db 3061 WTPAGASVPVVFITAVVLRDLADYKPERLLVHSAAGVGMBAVQALRHNGVHNGTA 3120
Qy 1121 GTPERRAYLESIGV--RVVSDSRDRFVADVRAMTGGEGVDVVLNSGELLIDKSFNLIR 1178
Db 3121 SHGKMA--LRALGLDHAIIASSRTLDFFESAFRAASGAGMDVVLNLSAREFVDSRLIG 3179
Qy 1179 SHGRFVELGKRCY--ADNOLGLRPLRLNLSFSLVDLRGMLEPRPAVRLBELGLIA 1237
Db 3180 PGRFPEWKGKTVDRODERVAADHPGVGYRAFDLGE-----AGPERIGMLLEVALLFED 3233
Qy 1238 GVFTPEPIATLPLARVADAFRSMAQHLGKLVLTG--DPEVORITPHAGA----- 1288
Db 3234 GYLRLHPVTTWVRARDAFRHVSQARHGKVLVTPSLDDEBGVLLVGGTALOGIYA 3293
Qy 1289 ----- 1288
Db 3294 RHVGEWGVRLILVSRGTDAPAGELVHELEAGADVSAACVADREALTAVLDSIP 3353
Qy 1289 ----- 1288
Db 3354 AEHPITAVVHTAGVLSGTLPSTMTADVEHLRPKVDAFLDELSTPGYDLAFAFMS 3413
Qy 1289 ----- 1288
Db 3414 SAAAVFGAGGAGYAAANATLDAAMRRRTAGLPALSLGWLMAETSGMTGLSDTRSR 3473
Qy 1289 ----- 1288
Db 3474 LARSGATPMDELSTLDAAMRRDDPALVPIALDVAALRAQRODGMALPLLSGLTRGR 3533
Qy 1289 -----GSTGDRDLDDLAAAPARAALAEPRTOVQVL--RTPEIKY 1332
Db 3534 VEGAPVNOGRRAAAGGGEADTDIGGLAAMTPDDRAHLRDVKTVAATVLTGHGTPS--RV 3592
Qy 1333 GAELFTRLGMDSLMAVELRNRIEASIKLSTTFLSTPSNTIALAQNLLDALATA 1388
Db 3593 DIERAFRDGFDLSLTVLELRNRLAATGLRLPATLVFDPHTPELAGHLDELATA 3648

```

## RESULT 10

T17409  
 polyketide synthase type I - Streptomyces venezuelae  
 C:Species: Streptomyces venezuelae  
 C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #ext\_change 17-Nov-2000  
 C:Accession: T17409  
 R:Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D. H.

```

Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A:Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A:Reference number: 218773; MUID:9644533; PMID:9770448
A:Accession: T17409
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4613 <XUE>
A:Cross-references: EMBL:AF079138; NID:g3808326; PID:g3808034; PIDN:AA69329.1
A:Genes: pikA1
C:Superfamily: acyl carrier protein homology
C:Keywords: antibiotic biosynthesis; carrier protein
P:1010-1081/Domain: acyl carrier protein homology <ACPI>
F:2495-2566/Domain: acyl carrier protein homology <ACP2>
F:4407-4478/Domain: acyl carrier protein homology <ACP3>

Query Match 28.5%; Score 2051.5; DB 2; Length 4613;
Best Local Similarity 34.6%; Pred. No. 2.8e-102;
Matches 589; Conservative 183; Mismatches 515; Indels 417; Gaps 52;

Qy 9 7AEDPIATVGAACRLPGGVLDLSGFWTLEGRDITVGRVPAER-WDAAMPDPDPAQX 67
Db 2601 ADDPIATVANSCTPIGDIRSPEDLMRLMSBEGGITTPTDRGMDLGLYADDAICR 2660
Qy 68 TPVTRASFSDVACDFDASFFGISPREALRMDPAHRLLEVCWALENAIAPSALVTET 127
Db 2661 AVVREGFLHDAEEDAEFFGVSPREALMDPQGMILLTTSWEAFERAGIEPALSRSST 2720
Qy 128 GVFTIGFSEYEAALPQATASAEIDAAGL--GTPPSGAGISIALGIRGCVAVDY 185
Db 2721 GVFTIGFSDYVAARPNAPRGVE---GYLLGSTPSVASGRITATFGEGPATVVDAC 2776
Qy 186 SSILVAVHACOSLSSGCSFTALAGVSLMSPSLTVMLSKTRALARDGCAFSABAD 245
Db 2777 SSLLTALHVAVALSGECTYALAGVAMATPHMFVEFSKRLAPDRSGAFSADAG 2836
Qy 246 FGRGEGCAVVVYKRLSGARADDRILAVYRGSAINHDGSSGLTVPNGSSQELVKRALA 305
Db 2837 FGAAGVGILLVERISDARRNHPVLAVVGRVAVQDGASNLTPNPGSQGRVIRQALA 2896
Qy 306 DAGCAASSGVYEAHGTTTACDPIEIQALNAVGLGDVATPLLIGSYKTNLGHPEVAS 365
Db 2897 DARLAPGIDAVETGTGTSIDPIEAOGLQATYKKEKPAERPLAIGSVKSNIGHTQAAA 2956
Qy 366 GTGLKLVLSHQOIPAHILQAOLNPRISGDLRLTYTRARTPMDNTRRARGVSF 425
Db 2957 GAAGITKVLAMRHGTLPTTLHADPSPAVDANSGLALVTEPIDMPAGTGRRAAVSF 3016
Qy 426 GMSGTNAVLEEAR-----AATCPAP-E 450
Db 3017 GISGTNAVYLEQADDAAGEVLEGADEVPEVSTVAMAGTSEVAESESSEAPAAPGS 3076
Qy 451 RPAEL-----LVLSARTASALDQAARLRDLHETVPSOC-----LGDVAFSLATRS 497
Db 3077 REASLPGHLPLWVLSAKDEQSLEGAALAHAMVSEPAADLSDAGPARLRDVTYLTATSKT 3136
Qy 498 AMEHLTAVAATSRREGRLAALDAAGOTSPPGAVRSIADSSR--GKLAFLFTGGAGATLGMG 556
Db 3137 AFARHPAAVTAARDPDLGLATLAGGTS--AHVHLDTARGGTTAFLTGGGSGSPGAG 3193
Qy 557 RGLYDVWSAFREAPF-LCYRLFNOELDRPLREVMMAEPASVDAALLDQTAFTQPALFTFE 615
Db 3194 RELYORHPFARALDEICHL--DGLHELPILDMVFRASGSAALLDERRYQCALFAL 3252
Qy 616 YALALMNSWGVPEPLVAGHSTIGELVAAACVAGVSELEDAVFLVVAARGLMQLPRGGMV 675
Db 3253 VALFRLVSWGMRPALILGHSVGEITAAHVAVGVFSLADAARLVAAARGRLMDLPAAGMIL 3312
Qy 676 STEAEADVAAAVAP---HAASVSTIAVANPQOVVIAGAGQGVNHAIAAMARGARTKL 732
Db 3313 AVQAADEIRVWLETREIRAGRLDVAAVNGPAAVLSGDADAAREAEVWSLGRRTRL 3372
Qy 733 HVSHAHPFLMAPMLDAFGRVASVSYRRPSITLVLSNLSGKAC--TDEVSSPGYVWRHARE 791

```



Db	3373	RVSAFAHSAHMDGMLDGFRAVLDEYEFRRPRLTIVSVNTGLAAGPDDLCDPEYVWYKRG	3432
Qy	792	VVRPADGVKALHAAGAGTVEVEGPKSTLLGLV-----PACMP----	828
Db	3433	TVRFIDGVRVRLDGLVRFCTCLEIGPQGVLTAAADGLADTPADSAAGSPVSGPASPADSA	3492
Qy	829	----DARALLASSRAGDEPATVLEALGGLMAVGLVSMAGLFP--SGGRVPLPTYWQ	883
Db	3493	AGALRPRLLVALLRKRSSETETADALGRAHAAGTGPDMAWPAWGSAHVVDPPTYSEFR	3552
Qy	884	RERYWIDPKADDAAGDRRAPGAGCHDEVEEGAVGCGDRSARLDHP-----PPESGR	936
Db	3553	RDRVWLDAPADTA-----VDTHGLGLG-----TDHPLGLGAVSLPD---3590	
Qy	937	REKVEAAGDRPRPLEI--DEPGVLDHLVLRVTERRAPGIAGEVIAVDA--GLSFNDVOLAL	994
Db	3591	RDGLLITLT----RLSTRHPWLADHAVLGVSV--LLPGAMWELAHAAESAGLDVDELT	3644
Qy	995	GMVPPDPLP-----GKPNPRLLLGCGCACRIAYAGCGVAVGLVVGQPVIALS	1039
Db	3645	LLEPLVLDBEGHGVLEKRVTVGAPAGER-----GGESEA-----GDGARPVLSHSLADAP	3692
Qy	1040	AG-AFATHVTTSAAVLVLRPPQALSAIEAAM-----PVAYLTAW	1077
Db	3693	AGTAMWSCHAT--GLLATDRPELVPVAPDRAAWMPQGAEEVPLDGLYERLDGNGLAFGPLF	3750
Qy	1078	YADRIATLQGE--RVLIHAAT-----GGVGLANQMAOH-----1111	
Db	3751	QGLNAWRYE--GEVFPADIALPATWTATAPATANGGSAAPAPYGIHPALLDASLHAIAVG	3809
Qy	1112	-----VGAEVHATAGTPEKRAYLESIGRVYVDSRSDRFVDVAPMTGCE	1156
Db	3810	GLVDEPELVVRPFHMSGVTVHA--AGAAARVRLASAGTDVAVSLTLD-----GE	3857
Qy	1157	GVDVVV-----LNSLSGELLIDKSFNL-----LRSHGRFVELGK	1188
Db	3858	GRPLVSVRLTLRPVTDQAASRVGGLMHVRAMPVYALASSGEODPHATSYGPTAVLIGK	3917
Qy	1189	RDC-----YADNOLGIKRPFLRNLSFSLVLDRCGMLEPRPARVALLBEELGLIAAGFT	1241
Db	3918	DELKVAALLESAGVEVGLYPDLLAALSOD-----VAAG--A	3950
Qy	1242	PPP--IATLP-----IARVADAFRSMAOAHG--KLVLTLG--DEPV	1278
Db	3951	PAPRTVLALPLPAGPADGAGEGVGRVARTLELQGMWLADHELACTRLLLVYRGAVRDE-	4009
Qy	1279	QIRIPTHAGAGBSTGDRPL-----LDRLASAPAPAPAAAL-----E	1314
Db	4010	-----GSGADADGGEDLSHAAMGLVTRTQTEHPGRFGILLDLADDASSYRTPSVLSD	4061
Qy	1315	AFLRTOVSQVLR-----TPRELKVGAEA-----1336	
Db	4062	AGLRDEPOLALHDGTIRLARLASVREPEGTAPALAPRGTVLLTGCGGGLGVARHVYG	4121
Qy	1337	-----LFTRLGDSLMAVELRNRIEASIKLKLSTFLSTSPNIALLAONLLDALATA	1388
Db	4122	EMGVRRLLLVSRRGTDAPGADDELVHELFA-----LGADVSAACDVADREALT	4169
Qy	1389	LSLERYVAEN-----LRAGVQND	1406
Db	4170	AVLDAIPEHEPLTAVVHTAGVLSD	4193

Nature 393, 537-544, 1998  
 A.Authors: Sqares, R., Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.  
 A.Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
 A.Reference number: A70500; MUID:96295987; PMID:9634230  
 A.Accession: H70621  
 A.Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A.Molecule type: DNA  
 A.Residues: 1-2126 <COL>  
 A.Cross-references: GB:285982; GB:ML123456; NID:g3261718; PIDN:CAB06632.1; PID:g1839013  
 A.Experimental source: Strain H37Rv  
 C.Genetics:  
 A.Gene: pks7  
 C.Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I h  
 nase homology; [acyl-carrier-protein] S-malonyltransferase homology  
 C.Keywords: carrier protein  
 F:56-456/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>  
 F:563-845/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AM11>  
 F:1694-1875/Domain: short-chain alcohol dehydrogenase homology <SADh>  
 F:1976-2051/Domain: acyl carrier protein homology <ACPP>

Query Match	Score	DB 2	Length
28.1%	2025	2126	

Matches 605; Conservative 184; Mismatches 567; Indels 734; Gaps 38;

QY 3 DRPIERAEDPIAIVGASCRLPGVIDLSGFWTLLEGSRTVGRVPAER-WDAAWFDPD 61

Db 26 NRDLARTTEPVAVVGMCRYPGGVDSPETLWELVAHGRDAVSEFPADRGWDVAGLFDPD 85

QY 62 PDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENAIAPSA 12

Db 86 PDAVGKSYTRCGFLTDVAGFDAEFFGIAIPSEALLMDPQQRLLLEVSWEALERAGIDPIT 14

122 LVGTETGVFIGIPSEYEALPQATASAEIDAHGGLTMPVGVAGRISYALGLRGPVAV 18

Db 146 LRGSGTGVFAGVFHGSYGG--QGRVPGLDERYGLRGSTLSVASGRVAYVLGLQGPAVS 20

QY 182 DTAYSSSLVAVHLACQSLRSGECTALAGVSLMLSPSTLVWLSKTRALARDGRCKAFSA 24

Db 203 DTACSSSLVALHLAVQSLRLGECDLALVGGVTVMATPAMFIEFSRQRLSADGRCKAYAG 26

242 EADGFGREGCAVVLKRLSGARADGRIILAVIRGSAINHGDASSGLTPNGSSQEI VLK 300

Db 263 AADGTAFAEGAGVLLARLADARRLGHPVLALVRGSAVNQDGSNGLATPNGPAQQRVIT 32

QY 302 RALADAGCASSVGYEAHGTTLGDDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 36

Db 323 AALASARLGVADVVEGHGTTLGDPRIEAQAILATYG-QRPADRPWLGSIKSNIHT 38

362 EYASGITGLKVL<sup>Q</sup>HGQI PAHLHAQALNPRI<sup>W</sup>GDRLTVTRARTPWPDPWNT<sup>P</sup>RRAG 42

Db 382 SAAGVAGVIKMVQAMRHGVLPKTLHVDVPTPHVDWSAGAVSLTEPRPWHVPCRPRAAG 44

QY 422 VSSFGMSGTNAHVLEAPATCTPPA--PERPAELLVSARTASALDAQARLRDHIET 47

Db . 442 VSSFGISGTNAHVILEEAPAVEPVGAHGNBPVAVPWLARSQAQALTNQARRLLAWVGA 500

QY 480 YPSQCLGDVAFSLATTRSAMEHRLAVAATSREGLRALDAAQQTSPGAVRSIADSSRG 53

Db 502 DENVRPLDVGWSLVNTRSLFDHRAVVVGADRTQLMEGLTGLAAGVPGADVAGRAQTV-G 56

540 KLAFLFTGGAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVWMAEPASVDAA 599

Db 561 KTAFFVFPQGQAWLGMGAQLCATAPVFAEHIHRCERALREHVESLLDVLGAPG---AP 61

600 LLDQTAFTQPALTFEYALALWRSWGEPELVAGHSIGELVAACVAGFSLDVAFLVA 655

Db 618 GLDRVDVQPALWAVMSLAELWRSVGVPDAVIGHSGEIAAAAYVAGALSLRDAAVVA 67

660 ARGRLMQALPAGGAMVSI EAPEADVAAA VAPHAASVSI AAVNAPDQVVIAGAGQPVHAIA 71

Db 678 LRSRLVRLGAGMVSACGQPAEKLASQWGDRLNIAAVNGVSSVVLGETDAVTELM 73

QY 720 AAMARGARTKALHVSHAFHSPIMAPMLEAFGRVAESVSYRRPSIVLVNSGKACTDEV 77

```

Db 728 ORCEAGRIARRIDVDYASHSAQVDAIREELIAALRGIEPRSTVAFFSTVIGELMDTAG 797
Qy 750 SSGPVWRRHAREVREPADGVKALHAAGAGTFVEVGPKSTLLGLVPAICMPD-----ARPA 833
Db 798 VNAEWTYSIRPQVPERAVRNAFDGQYAVFESSHPILLINGIETLVDCCRAGTGEPI 857
Qy 834 ILASSRAGDEPATVLEALGGLMAYVGL-----VSMAGLPPS--GGRVPLPTYPMORE 885
Db 858 VLPF--LGRDD-----GGVGRFMLSQAHVAGVDMRAAFADGGRVELPTAFAFANQ 910
Qy 886 RYVID----- 890
Db 911 RFWLMDGLAVGGDLGGVGLVGAHEGLLAAVORPDGGVLTGRISSVVAAPWLADHAYGP 970
Qy 891 -----TKADD----- 895
Db 971 VVLPFGTGVELALRAGDEVGCSVLQELTLQAPLVLPADGVAVQVVGVGSEOSTRANWV 1030
Qy 896 -----AARG----- 899
Db 1031 YSAAGADSSPGMTLHAQVGLGVGSQPAELSSVMPVGAAMDADYQVLAARGCYG 1090
Qy 900 ----- 899
Db 1091 PAFRGLOALMRGAEVADVTLPEGVPIRGFIHPAVIDDALHAWGIVEGQOTMLPFSW 1150
Qy 900 -----DRAP--GAGHDEVE----- 912
Db 1151 QGVCHASGAARVVRRLAVRGAVSVELADQGLETVSRQDMVRPVSAALSTAGD 1210
Qy 913 -----EG-----GAVRGD----- 921
Db 1211 RGLLEMTVPVLEGGDIDGDAVWMLPRHAGAQGDVLAAYVGVHEVLEVLSWLAS 1270
Qy 922 -----RSARLDHP----- 930
Db 1271 DATGLVVTTRGAVGPVDDVTLDAAGAAVWGLVRSQAHEHPRVVLVDTGSAVEDAVG 1330
Qy 931 -----PRESGRREKVEAAGDR-----FRLIDERGVLVDHLVLTERR 969
Db 1331 FGARGEPQLVVRKGVVTAARLAPVAAGLTPRSASAGKRLVAGGGGTADVAVPVAV 1390
Qy 970 APLGEVEIADVDAAGLSFNDVOLALGMVBDLPKRNPLLLGSCAGRIYAVGSGVGL 1029
Db 1391 ELATQVRAVYAGVGNFRDVLVALGMYPG--GE-----LGVGAGVVEVGCVTGL 1442
Qy 1030 VVGQPVIALSAGAFTHVTTSAALVLRPQALSIAEAAAMPVAYLITAMVALDRIRLOPG 1089
Db 1443 AVGDVWML--LGLVSEAVVDARLVTMVPAGMSLVEAAVPAFLTAFGLSVLAEVAG 1501
Qy 1090 BRVLIHAATGGVGLAAVQMAHVGAEVHAHTACTPEKRAVLESLGR--VYSRSRDRFA 1147
Db 1502 QVVLHAGGVGMAVSLARWGAHVFTA--SRAMDTLRMGDDHIIISRSRLBEE 1560
Qy 1148 DVRAWTGEGVDVUNLSLSEGLIDKSFNLLRHGRFVELGKRD-----CYADNOGLRPF 1202
Db 1561 AFLRTBESSGVUVVUNLSLAGEFTDASLRLLPBGGRFIELGKTDIDGQVVAERHNG-- 1616
Qy 1203 LKNLSFSLVDLRGMMLEBPAPVALLLEILLGIIAAGVTPPPIATLPIARVADAFRSMQ 1262
Db 1617 VRYRAFIDIVE-----AGPDRIAAMLSEVGLLAAGVLAARLPVKTFDARCAABAARFVSQ 1670
Qy 1263 AOHGLKIVLT----- 1273
Db 1671 AKHIGKVVLTIPDGGGOSGLAGTVVVTGTGMAGSAVATHLVRRHGVANLVLSRSGE 1730
Qy 1274 -----GDPEVOI-----RTP----- 1283
Db 1731 QADRAEVAALIRBEGAQVAVVSCVADRDALLALLAGLDPVPLKGVFHAAGVLDAYI 1790
Qy 1284 ----- 1283

```

```

Db 1791 TGLPFRVDVTLRAKVDGAMNMLHETLMDLSAFVFPSSMAGIVGTPAQGVYAAANAFLD 1850
Qy 1284 -----THAGAPSTGDRDL--LD--- 1299
Db 1851 GLVAYRBRGLAGLSVAMGLMEQASAMTRHLGERDRAAMTQAGLAPLTTEQALGLDRLAL 1910
Qy 1300 -----RLASAPARAALAEAFLR-----TQVSO---VLR---T 1327
Db 1911 QADRAVVAARLDRALAGAGALPALPSQLAAGPTRRIRIDAADTAVSMGLVSRHALT 1970
Qy 1328 PE-----IKVGAELFRLQMDSLMAVELNRRIRIASLKLALS 1364
Db 1971 PERQRELTDVVISMAAVLGRSSSVVDINAKAFODLGFDSLTVAVELNRRIRKTAATGLTIS 2030
Qy 1365 TFLSTSPNIALAQNLLDALATALSERVAENLRACVQNDVFSSGADQ 1414
Db 2031 PTLIFDYTPPATLAEHLDSRLVTA-----SSSDQ 2059

RESULT 12
A70984
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jul-2000
C:Accession: A70984
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70984
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2188 <COL>
A:Cross-references: GB:263857; GB:AL123456; NID:G3242252; PIDN:CAB06099.1; PID:G3242253
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: ppac
C:Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hc
nase homology; [acyl-carrier-protein] S-malonyltransferase-homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:56-460/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1482-853/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F:1482-1770/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:1802-1982/Domain: short-chain alcohol dehydrogenase homology <ACPI>
F:2072-2141/Domain: acyl carrier protein homology <ACPI>
F:2105/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 27.9%; Score 2014.5; DB 2; Length 2188;
Beet Local Similarity 31.7%; Pred. No. 1.1e-100;
Matches 565; Conservative 172; Mismatches 511; Indels 537; Gaps 35;
7 ERAEDPIAIVGACRLEPGYIDLSGFTLLGSGRDTGVRPAERMDAAAFDPDPDAPG 66
Db 30 EKSSSEPIAIVGACRLEPGYIDLSGFTLLGSGRDTGVRPAERMDAAAFDPDPDAPG 89
Qy 67 KTVPTRASFLSDVAC--FDASFPGISPREALRMDPAHRLLEVCWALENAIAPSAVLG 124
Db 90 TICSGEGFLTSWQDEDFDAEFFSISPREAAMPQOGLLLEVALEADAGVPQHTIRG 149
Qy 125 TETGVFIGISGEYEAALPQATASAEIDAHGGLGTMPSVGVGGRISYALGAGPCVAVUTA 184
Db 150 TQTSVFGVTAIVDMVLTLAGRLRPVDLDAYIPTGNSAFAGRLAYIIGARPAVVIDTA 209
Qy 185 YSSSVAVHACOSLRSGECSTALAGVSLMISPTLWLSKTRALARDGRCKAFSABAD 244
Db 210 CSSSVAVHACOSLRSGESDMALVCGTNLLSLGSPSIACGRWMLSPBEGCKTFDABAD 269
Qy 245 GFGRGECVAVVVKRLSGARDGDRILAVITGSAIINHDSAGSLGTVPGSSQETVLKRAL 304
Db 270 GYVRGEGAAVVVVKRLSDAVDGNRIIAVVVGSAVNQDGAASSGVTVPNGPAQOALLAKAL 329

```



Db 3336 HGRVNGYIG-----AGADIGFPAATASATSVLSGRVYFEGLEGPAITVDYAC 3383  
 QY 186 SSSLYAVHLACOSLRSCEGCTALAGVSLMLSPSTLWMLSKTRALARDRCFAASADG 245  
 Db 3384 SSSLYAVHLACOSLRSCEGCTALAGVSLMLSPSTLWMLSKTRALARDRCFAASADG 3443  
 QY 246 FERGGCAVVLKRLSGADADGRIILAVIRGSAIINHGAASSGLTVNGSSQETVLKRALA 305  
 Db 3444 TMAEVEGVTLVERLSDAQAKGHQVLAVERSSAVNDDGASNGLTANPNSQORVIOALIS 3503  
 QY 306 DAGCAASSVGYEAHGTCTTGDPIEIOALNAVYGLGRVATPILIGSKYKTLGHEVAS 365  
 Db 3504 NGGLAAHEVDVVEAHGTCTTGDPIEIOALNAVYGLGRVATPILIGSKYKTLGHEVAS 3561  
 QY 366 GTGLGLKVLVLSOHGOIPALHQAOLNPRIISWCDLTLVTFRATPMPDNTPRRAGVSSF 425  
 Db 3562 GVSQVTKMVALOHNTVPRTLHVDEPSRVDNAAGAVELRENOQPMFGTDRPRRAGVSSF 3621  
 QY 426 GMSGTNAHVLEBPATCTPPAPE--RPAEL-----LVLSARTASALDAQAARLMD 475  
 Db 3622 GVSQTAHVILIESAPPA---QPAEBAQPVETTPVAVSDVLPVISAQTOFALTEHEDRLA 3678  
 QY 476 HETTPSQCLGDVAFSLATTRAMEHRLAVATSRGLPAALDAAQCO--TSPCAVRSI 533  
 Db 3679 YLAASPGADIRAVASTLATVRSVFEHRAVLG-----DDTVGTAVTDP----- 3722  
 QY 534 ADSNRKLAFLFTGQGAOTLGMGRGLYDVMSAFREAFDLCVRLFNQELDRPLREVMMAP 593  
 Db 3723 -----RIYFVFPQGMQWMLGMSALRDSVVFPAEMACAAALREFVWMDLFTVL----- 3772  
 QY 594 ASVDAALLDQTAFTOPALFTFEYALALWRSWGEVELVAHSGISGLVAACVAVGESLED 653  
 Db 3773 --DPAVVDVAVVQPSAWAMVSLAAVQAGVREDAVIGHSQGIIAACVAVGESLED 3830  
 QY 654 AIFVLAARER-LMOLPAGAMVSTAEPAADVAANAAPAAVSTIAVAPDQVYIAGG 712  
 Db 3831 AARIYTLRQALARGAGGAAVVALPQADVEL-----VDGAMIIAHNGPASTVYIAGP 3885  
 QY 713 QPVHAIAMAAMARGARTKALHVSHPSPMLMAPLEAFGRVAESVYRSPSTVLVNSLGG 772  
 Db 3886 EAVDVLTLHEAGVAVRRIYDYASHTPHVELIRDELDTISDSSQPLVPMJSTVGG 3945  
 QY 773 KACTDEVSSPGYVVRHAREVVRPADGVKALHAAAGCTFVEVPGKSTLLGLVPACMPDARP 832  
 Db 3946 -TWVSPDLGEYWRNLRBPVGFHPAVSQLOAGDVFVFSASPVLLQAM-----DDVV 3999  
 QY 833 ALLASSRACRDEPATYLEALGGMAGVLSWAGLFPSGGRRY-PLPTYPKQERYWIDT 891  
 Db 4000 VTVATLRRDDGDATRMILTALAQYVHVGVTVDPWPAIIGTTTTRVLDPFYAFQHQRYWLS 4059  
 QY 892 -----KADDAARGRRAPGAGH----- 908  
 Db 4060 VDRAAADGHPILGTVVALLPESGCVLITGRVSLATHTWLADHVAHSVLLPGTAFVELVVR 4119  
 QY 909 --DEVE----- 912  
 Db 4120 AADEVCEVDVDELVIETPLLLPQTGGVQLSVSAEADSGHRTVTVFSQADNTDAIRHV 4179  
 QY 913 -----EG-----GAVRGCD-- 921  
 Db 4180 SATISTSTPLSPESDTPWPAQAPMNVADFYDRLMAGYEXGAPQGLQAMRGDGV 4239  
 QY 922 -----RSARLD--HPP-----PSSGRREYVEAGD-- 945  
 Db 4240 YAEVALAEQAOBAARYDVHAPLLDALHACTLNTPDAEBGVGLPSSMNVQVHAATPAT 4299  
 QY 946 ----- 945  
 Db 4300 LRAVATQBADGWSLRVADDTGRVATIGSLITRPVATDLTGSTTDDLLTLTWTVIPRPOE 4359  
 QY 946 -----RPR 949  
 Db 4360 TSLPMVRYEDLADGNMPVDPVVVFTALPNSSENPLGPRPADPDLAOPDTDRQVLPKPO 4419

QY 950 -----LE 951  
 Db 4420 TWLPCTFNTDSTLVARTGTGLAAANVSGLMRGAQSHPRFLLVESDDDTLTLRQLAATVG 4479  
 QY 952 IDER-----GVLDHLVLRTERRA 970  
 Db 4480 LDEPRIRINDGREYEAARLARTRTTPEDTPTLTIPDSRAWLLEOPRGSTLODILLVPTDPAE 4539  
 QY 971 PGL--GEVEIADVADAGLSFNDVQALGNVPDPLPKPMPPLLLGGECAGRIYAVEGVNG 1028  
 Db 4540 QPLRPEVRIDVRAAGLFRDVLVAGVNDRP-----TGGEAAGVLEVEGPDVD 4591  
 QY 1029 LVVQGPVIALSAGAFATVTTTSAALVLRPPQALSAIEAAMPVAVLITMAYALDRIARLP 1088  
 Db 4592 LARGRVGVGVGSGSPRAALDRMLAVIPDQMSSTTAASVAVPATVATVYGLVDLAGLSE 4651  
 QY 1089 GERVLIIHAATGGVGLAAVQMAOHVGAEVHATATPEKRAYLES-LGVRYVSDRSRDPVA 1147  
 Db 4652 GESVLIIHAAGGVGAATQIARHLGARIVATSTGKHILREAGLEDTHIADSRILSFOE 4711  
 QY 1148 DVRAVTGEGVDVUNLSLSELIDKSFNLSHGRFVELGRDCYADNQLG-REPLRLV 1206  
 Db 4712 TFLNTHQOGVDVUNLSLSDGFVDASLDLPRGGRFVEMGKTDIRDPHQVADRFGTTYQ 4771  
 QY 1207 SFLVDLRGMLEPRARVALLLEELGLIAGVFTPPPIATLPRIARVADAFRSMQAQHL 1266  
 Db 4772 AFDLMD-----AGDRILREITTELLTLFTQGVLLPLVQAMDIQARDAFSWMGRARI 4825  
 QY 1267 GKVLVTL--GDPEVOIRIPTHAGPSTGDRDLDRLASAARPAALAEFLFTVQSO 1323  
 Db 4826 GKIVITPRRPPDDTILIT--GSGSVLAG--ILARHAAEHGARHLL--LSRTTPDQ 4878  
 QY 1324 VARTPEIVGAALTRGMSIMAVELNRLEASLXLKSTFTSTPNIALLAQNLLD 1383  
 Db 4879 ALIKELAEIGAH-----VDATC-----DVSDRGLARVLAGVSEPHPLTA----- 4919  
 QY 1384 ALATALSLEVAENLRAGVQNDPVSQADQDMEI 1418  
 Db 4920 VHTHAGALDDGVESLTTQODTVLRPKADGMHL 4954

## RESULT 14.

T17463  
 r1famycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei  
 C1:Species: Amycolatopsis mediterranei  
 C1:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 17-Nov-2000  
 C1:Accession: T17463  
 R1:Schupp, T.  
 submitted to the EMBL Data Library, December 1997  
 A1:Reference number: Z18802  
 A1:Accession: T17463  
 A1:Status: preliminary; translated from GB/EMBL/DBJ  
 A1:Molecule type: DNA  
 A1:Residues: 1-4735 <SCH>  
 A1:Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CA11035.1  
 A1:Experimental source: strain LBG A3136  
 C1:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology  
 C1:Keywords: carrier protein  
 F1:53-500/Domain: acetate-CoA ligase homology <ACU>  
 F1:543-610/Domain: acyl carrier protein homology <ACP1>  
 F1:2102-2173/Domain: acyl carrier protein homology <ACP2>  
 F1:3078-3150/Domain: acyl carrier protein homology <ACP3>  
 F1:578-4649/Domain: acyl carrier protein homology <ACP4>

## Query Match

Query 27.8%; Score 2002.5; DB 2; Length 4735;  
 Best Local Similarity 37.9%; Pred. No. 1.3e-99;  
 Matches 586; Conservative 183; Mismatches 526; Indels 251; Gaps 50;

QY 2 ADPRIE---RAEDPIATVAGASCRLPGVYIDLSGWTLLBGRDVTGVGVPAR-WDAAM 57  
 Db 2182 ANRPAEIGTAAEEPIAIVAACRPFGVHSPEDLMRLVADGADAVTEFPADRGWDTDR 2241



```

Db 3318 DAAGSVABEGGLDTAAGFDAGFFGISPREALANDPOORILLLEVSWAEFERAGIEPGSV 3377
Qy 123 VGTETGVIGIGPSEYEALPOATASAEIDAHGKLG--TMPGVGAGRISYALGRLGPCVA 180
Db 3378 RGSDDGVITGAFPVGYGAGA-----AREGIGATPAFVNLGRLSYFFGLGCPAIT 3427
Qy 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMSPSTLWMLKTRALARDGCKAFS 240
Db 3428 MDTACSSSLVHLALHAAQALRNGECMALAGVTVATPEVFTEFARQGLADGCKAFS 3487
Qy 241 AEAODFGGECVAVVVKRLSGARADGRIILVINGSAIHDGASSGLTPVNGSQEIVL 300
Db 3488 DSADAGSEBEGGLLVERLSIPARRNGHQLAVVGSANQDASNGFPAFPAQOQRYI 3547
Qy 301 KRALLADACCAASVGVVEAHGTGTLGPDIETQALNAVYGLGRDVALPLLSGVNTNGH 360
Db 3548 RQALNAGLTTRREVUNVEAHGTTLGDPIEAOYIATYGORE--QPLLDGTLSSNGH 3605
Qy 361 PEYASGITGLKVLVSLQHQOI PAHLHQAALNPRI SWGDLRLTVTRARTPMDMNTPRRA 420
Db 3606 TQAAAGVGVIMVVALQHSITVPRTLHVNEPSRHVMSGAVELVTENQSWVTGRPRRA 3665
Qy 421 GVSAGSGMGTNAHVLEAPAA---TCTPPA-PRPAPL--LVISARTASALDQAARL 473
Db 3666 GVSAGVSGTNAHVLEAPAAQSVNNAPVATPVVASLVLVISAKTLPALTEHEBRL 3725
Qy 474 RDHETVPSQCLGDVAFSLATTRSAMERHLAVALTSREGLRAALDAAQGGSPGAVRSI 533
Db 3726 RAYTLASGADMRVAVGSTLALTRSVFERN-----AVLLGHDTVTVTGCG 3769
Qy 534 ADSSRGKLAFLFTGGAGQTLGMRGLYDWSAFREAFDLCVRLFNOELDRPREVMAABP 593
Db 3770 TAVSNPRVVFVPGQGMQWLGMSALRSGSVFAFARMAECAAALBEPFMDLFAVL----- 3825
Qy 594 ASVDAALDDQRTFOTPALFTFEYALALMRSGVBEPELVAGHSIGELVAAACVAGFSLD 653
Db 3826 --DDPAVVDRAVDVQPPASMAVWVSLAAVWQADGVPRDVTIGHSGEIIAAACVAGVSLRD 3883
Qy 654 AVFLVAARGR-LMQALPAGGAMVSI EAPDVAALVAAPHAASVSI AAVNAPDOVVIAGNG 712
Db 3884 AARSTTLRSQAIARGLAGRNAASVALPAHEIEL-----VDGAMIAAHNGPSTVAGAP 3938
Qy 713 QPVHAIAMAARARARTALHVSASHFSPMLAPMLERFGRVAESVSYPSPISVLVSLNG 772
Db 3939 EAVDRVLVHEARGVVRIRIANDVASHTRPHELIDELLDITAGIGSQAPVPMWSTYDQ 3998
Qy 773 KACTDEVSP--GTVWPHAREVVRPADGVKALHHAAGCTFVEVPKSTLLGLVPACMPD 829
Db 3999 ---TWVEGFLDVEYWRNLREPVGDSAVQGLRAEGDVFVEVGSASPVLLQAM-----D 4049
Qy 830 ARPALLASSRAGRDPATVLEALGGLMAVGLVSWAGLFPSGGRVPP--LPTVPMQRERYW 888
Db 4050 DDVVTVAITLRDDGATRLTLALQAFVEGVTVDPALIGTATTVTRVPDLPTAFQHORFW 4109
Qy 889 IDTK-----ADDAARCDRRAPAGH--- 908
Db 4110 AEGADRSVAGHPPLGLVAVELPESDGVLLTGRVSLATHAMLADHAVRGSVLLPOTGFVBL 4169
Qy 909 ---DEV----- 911
Db 4170 VVRADEVGCDVDELVIETPLLLPSASVHLSVSGEADSGRRGVTVFSRADGADAMT 4229
Qy 912 -----BEGGAVRG----- 919
Db 4230 RHVSATIGVGAALSLPELAMPPAQAOPVGLGDFYDRLTGAGYEGYGPAGLOQAMRDG 4289
Qy 920 ----- 919
Db 4290 DTVFAEVALAEQAEAAAFVHPALDLDAALDAGITLNTLDNAEQGVRLPFSMNGVQVART 4349
Qy 920 ----- 919
Db 4350 GSATLRVADNPTSDGMSVRAVDOSGRPVATVDSLVTTRPVTAADTLGSAADLLTVVWTEIP 4409

```

```

Qy 920 ----- 919
Db 4410 TPQOTGLSVGRFEDLADGDVPEVVVCTALPDSESNELAPLDPDPLVQTRTLTTOVLQ 4469
Qy 920 -----GDR-----RSARLDHP----- 930
Db 4470 AVQAWLAGERFTDSTLVTRGTGLATAGVSGLMRSASQSEHGRFVLVECDNLTLOQLAA 4529
Qy 931 -----PPSGRREKVEAA-----GDPFFLEIDEPEVLDHLVL--RVT 966
Db 4530 TVGLDEPRLVCDGRFEVPRRLARANTPESSSLPTTIPGDAMULEQSHSGTLRLDALVPAET 4589
Qy 967 ERRAPGLGEVEIAVDAAGLSFNDVOLALGMPDDLPGRKPNPRLLLGSCAGRIVAVGCV 1026
Db 4590 AERPLQSGEVAVDRRAAGLNFRLDUALGTYPGF-----AVIGABAAGVLEVEGVEV 4641
Qy 1027 NGLVVQOPVIALSAGAFVTHVTTSAALVPRPPQALSAIEAAMPVAVYLTAMVALDRIARL 1086
Db 4642 QDLAPGRDVFGLVGGFGCAVALIADRRMLGVIPDGMSFTTAASVPVVFATAYYGLVDLAGL 4701
Qy 1087 QPGERVLIHAATGCVGLAAVQMAQHVGAENVATAGTPPKKAYLES-LGVRYVDSRSDRF 1145
Db 4702 SAGESVLIHAAAGGVMAATQIARHLGARIVATASGKHVLRXGLDARIGDSRTTG 4761
Qy 1146 VADVRAWTGEGGVDPVNLNSLGEILDKSFNLLRSHGRFVLEKGRDVCYADNOLGL-RPFLR 1204
Db 4782 REMULDTDSRGVDVNLNSLGSDFVDASDLPLPRGRFVEMGKTDIRPHOVYADRPGTS 4821
Qy 1205 NLSFSLVDLRGMLEPARVRLLEELIGLIAAGVFTPPPIATVPIARVADAFRSMAQAQ 1264
Db 4822 YQAFDLM-----AGPRLREIITADLLALFAQGVLLPLVRAMDIRQAREAFSWSRAR 4875
Qy 1265 HLGKVLVTLGPPEVOIRIPTAGAGPS---TGDRDLRLASAPAAALAEALFRTQV 1321
Db 4876 HIGKIVLT-----VPOOLDANGTVLVGSGSVL-----AGIAARHLVAEOGVRL 4921
Qy 1322 SOVLRTEPKVGAELFTRLGMSLMAVELNRIEASL-----KLKSTFLSTSPNAL 1376
Db 4922 LLSSTPD-----DALINELG-----ELGARVDIACVDSRAGLARIILAGVSPEHPL 4969
Qy 1377 LAQNLDALATLSLERVAENLRAGVQNDVSSGADQDWEI 1418
Db 4970 TA-----VIHTAGALDDGVESLTAQOLETVLRPRKADGAMHL 5006

```

Search completed: October 2, 2003, 17:35:08  
 Job time : 93.4261 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:34:03 ; Search time 81.0897 Seconds  
(without alignments)  
2772.494 Million cell updates/sec

Title: US-10-014-717-2  
Perfect score: 7210  
Sequence: 1 VADRPPIRAEDPIAIVGAS.....GVQNDVFSGADQDWEIHAL 1421

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCRUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7210	100.0	1421	14	US-10-014-717-2
2	4137	57.4	2439	14	US-10-014-717-7
3	3311.5	45.9	7257	14	US-10-014-717-7
4	2311.5	32.1	6146	15	US-10-156-761-10436
5	2249	31.2	6145	15	US-10-156-761-10436
6	2245.5	31.1	3798	14	US-10-014-717-6
7	2199.5	30.5	3352	15	US-10-156-761-7961
8	2197.5	30.5	3564	15	US-10-156-761-7964
9	2159.5	30.0	7746	15	US-10-156-761-7963
10	2152	29.8	1835	15	US-10-156-761-7963
11	2149.5	29.8	1784	15	US-10-205-032-18
12	2128	29.5	3816	11	US-09-808-880-2
13	2125	29.5	5215	10	US-09-861-289-2
14	2125	29.5	5215	10	US-09-860-846-2
15	2125	29.5	5215	11	US-09-988-384B-2

16	2125	29.5	5215	11	US-09-836-821-2	Sequence 2, App11
17	2114.5	29.3	3745	15	US-10-205-032-14	Sequence 14, App1
18	2105	29.2	3939	15	US-10-156-761-10434	Sequence 10434, A
19	2077.5	28.8	3613	15	US-10-156-761-10432	Sequence 10432, A
20	2052.5	28.5	3739	10	US-09-861-289-33	Sequence 33, App1
21	2052.5	28.5	3739	10	US-09-860-846-33	Sequence 33, App1
22	2052.5	28.5	3739	11	US-09-988-384B-33	Sequence 33, App1
23	2052.5	28.5	3739	11	US-09-836-821-33	Sequence 33, App1
24	2052.5	28.5	11877	10	US-09-861-289-6	Sequence 6, App11
25	2052.5	28.5	11877	10	US-09-860-846-6	Sequence 6, App11
26	2052.5	28.5	11877	11	US-09-836-821-6	Sequence 6, App11
27	2052.5	28.5	12199	11	US-09-988-384B-6	Sequence 6, App11
28	2051.5	28.5	4551	11	US-09-793-708-1	Sequence 1, App11
29	2051.5	28.5	4551	12	US-10-201-365-2	Sequence 2, App11
30	2051.5	28.5	4551	12	US-10-160-539-1	Sequence 1, App11
31	2051.5	28.5	4613	10	US-09-861-289-31	Sequence 31, App1
32	2051.5	28.5	4613	10	US-09-860-846-31	Sequence 31, App1
33	2051.5	28.5	4613	11	US-09-988-384B-31	Sequence 31, App1
34	2051.5	28.5	4613	11	US-09-836-821-31	Sequence 31, App1
35	2051.5	28.5	4840	15	US-10-156-761-10435	Sequence 10435, A
36	2048.5	28.4	3739	11	US-09-793-708-2	Sequence 2, App11
37	2048.5	28.4	3739	12	US-10-201-365-3	Sequence 3, App11
38	2048.5	28.4	3739	12	US-10-160-539-2	Sequence 2, App11
39	2046.5	28.4	4881	15	US-10-156-761-8481	Sequence 8481, App1
40	2025	28.1	1925	15	US-10-205-032-12	Sequence 12, App1
41	2024.5	27.3	3970	15	US-10-156-761-10429	Sequence 10429, A
42	1966.5	27.1	6396	12	US-09-940-316B-72	Sequence 72, App1
43	1954.5	27.1	1832	14	US-10-014-717-4	Sequence 4, App11
44	1918.5	26.6	4150	11	US-09-808-880-2	Sequence 2, App11
45	1889	26.2	3519	11	US-09-808-880-4	Sequence 4, App11

## ALIGNMENTS

RESULT 1  
US-10-014-717-2  
Sequence 2, Application US/10014717  
Publication No. US20020192778A1  
GENERAL INFORMATION:  
APPLICANT: Schnupp, Thomas  
APPLICANT: Ligot, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/10/014,717  
CURRENT FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US/09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1421  
ORGANISM: Sorangium cellulosum  
US-10-014-717-2  
Query Match 100.0%; Score 7210; DB 14; Length 1421;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VADRPPIRAEDPIAIVGASCRIPGVYIDISGFWTLGSRDVTYGVPAERMDAAAFDP 60  
DB 1 VADRPPIRAEDPIAIVGASCRIPGVYIDISGFWTLGSRDVTYGVPAERMDAAAFDP 60  
QY 61 DPDAKGTPTTRASFLSDVACFDASFGISPREALRMDPAHRLILEVCWEALENAATAPS 120  
DB 61 DPDAKGTPTTRASFLSDVACFDASFGISPREALRMDPAHRLILEVCWEALENAATAPS 120  
QY 121 ALVETETGEVIGISPEYEALPQATASAEIDAHGIGLTPSVGAGRIAYALGLRGCV 180

[illegible]

D	b		1201	FPLRLSLSYVDLNGMKMLERARPARALLLEGLIAGVFPPPIATLPJRVADAFRSM	1260
O	y		1261	AQOHGKLVLTLDGPVEVOIRIPFHAGAPSTGBDLDLRASAAPARAALAEFLRTQ	1322
D	b		1261	AOQHGLKVLVTLLGDPVEVOIRIPFHAGAPSTGBDLDLRASAAPARAALAEFLRTQ	1322
O	y		1331	VSOVTRTEIVGAELPTRIGMSLMAVELRNBIEASLKIKLSTTFPASTPNIALAQN	1386
D	b		1331	VSOVTRTEIVGAELPTRIGMSLMAVELRNBIEASLKIKLSTTFPASTPNIALAQN	1386
O	y		1381	LDDALATLSERVAAENIRAGVONDFVSSGADDMETIAL	1421
D	b		1381	LDDALATLSERVAAENIRAGVONDFVSSGADDMETIAL	1421
				RESULT 2	
				US-10-014-717-7	
				; Sequence 7, Application US/10014717	
				; Publication No. US20020192778A1	
				; GENERAL INFORMATION:	
				; APPLICANT: Schupp, Thomas	
				; APPLICANT: Ligon, James	
				; APPLICANT: Molnar, Istvan	
				; APPLICANT: Zirkle, Ross	
				; APPLICANT: Cyt, Devon	
				; APPLICANT: Goelach, Joern	
				; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOPHILONES	
				; FILE REFERENCE: 4-30582A	
				; CURRENT FILING DATE: 2001-11-13	
				; PRIOR APPLICATION NUMBER: US/10/014,717	
				; PRIOR FILING DATE: 1999-06-17	
				; NUMBER OF SEQ ID NOS: 30	
				; SOFTWARE: PatentIn Ver..2.0	
				; SEQ ID NO 7	
				; LENGTH: 2439	
				; TYPE: PRT	
				; ORGANISM: Sorangium cellulosum	
				US-10-014-717-7	
				Query Match	
				Best Local Similarity 57.4%; Score 4137; DB 14; Length 2439;	
				Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13.	
O	y		6	IERAAEDPIAVGASCRLPGVIDLSCFWTLLEGSRDVGVPAERMDAAMPDDPDAP	65
D	b		26	LEQERTETIAVGICRRPGGADTPBAFMWELDSGRADV--QLDRRALVGVHPSEEVF	83
O	y		66	GKPTVPTRASPLSD-VACFDASFEGISPREALRMDPAHRLLEVCEWEALENNAINPSALVG	124
D	b		84	-----RWAGLLTTEAVDGFDAFPFGTSPREARSIDPQORLLLEVMEGLEDGIAIQSIDG	138
O	y		125	TETGVFIGISEXEYEAALPQATAAEIDAHGCLGTMPBSVGRI SYVALGIRGPCVAVDTA	184
D	b		139	SRTGEFLDACSDDSHTVAAQQRR-EODAYDTTGNTLSVAAGRSITYLGLGPECLTYDTA	197
O	y		185	YSSSLVAHTLACOSLRSGECSTALAGVSIMLSPSTVLSTKTALARDGCKAFSAEAD	244
D	b		198	CSSSLVAHTLACRSIRARESDLALAGVMNLLSKTMIMLGRIQALSPDHGCRFFDASAN	257
O	y		245	GFGREGCAVVVLRKLSGARPADGDRIIAVIRGSAINHDCASSGLTPVNGSQEIYVKRAL	304
D	b		258	GFVRGEGCGMVVLRKLSDAQRHGDRIIMWLIRGSAMNQDGRSTGLMAPVLAQEAALLREAL	317
D	b		305	ADAGCASSVGVYEAHGGTTLGDPTELQAANAYYGGLRDVATVLIGSVYTKNHGHPYA	364
O	y		318	QSARVDAGAITYVBTHGISTGLGPITEVARALVGPARRADSGCVGAVENTNLGHLEGA	377
O	y		365	SGITGLKLVVLSLHGQIPAHHLAOLNPRIISMGDLRLTVTRARTPMDDMTPRRAGVSS	424
D	b		378	AGVAGLIAALALHHELIIRKLHHHTLNFRIRIGTALATALATEEVPRPRAGRPRFAGVSA	437
O	y		425	FGMSGTNAVVLIEAPRAACTTPPAERPDAELLVLSARTASALDAQAARLRDHLLETYSQC	484

```

Db 438 FGLSGTIVHVLLEAPATVLAAPATPGSAELLVLSASAAALDQAARLSAHIAAYEEOG 497
Qy 485 LGDVAFLAATTSRSMERHIAVAATSRREGIARALDAAAGCOTSPGAVSIADSSGKIAFL 544
Db 498 LGDVAFLVSTRSMERHIAVAATSRREGIARALDAAAGCOTPAAGRAASSGKIAFL 557
Qy 545 FTGOGAOTLGWGRGLYDWSAFREAFDLCVRLFNOELDRPLREYMAEPASVDALLDOT 604
Db 558 FAGGAGAPVPGMGRGLMEAMPARETFRCVTLLFDRELHQPLCEYMMMEPSSRSLLDOT 617
Qy 605 AFTOPALFTFEYALAAIWRSGVEPELVAGHSIGELVAACVAGVPSLEDVAVLAAGRL 664
Db 618 AFTOPALFALEYALAAIWRSGVEPELVAGHSIGELVAACVAGVPSLEDVAVLAAGRL 677
Qy 665 MQALPAGAMVSIAPPEADVAANAAPHASVSIANAAPDOVYAGAGOPHATIAAAMAA 724
Db 678 MQALPAGAMVSIAPPEADVAANAAPHASVSIANAAPDOVYAGAGOPHATIAAAMAA 737
Qy 725 RGARTKALHVSASHPSPLMAPELAFGRVAVESVRSPIVLVSNLSGACTDESSSPGY 784
Db 738 RGARTKPLHVSASHPSPLMDPELAFGRVAVESVRSPIVLVSNLSGACTDESSSPGY 797
Qy 785 WVRHAREVVPADGVKALHAAAGCTFVEVGPSTLGLVPAQMPDARPAIIASSRAGDE 844
Db 798 WVRHAREVVPADGVKALHAAAGCTFVEVGPSTLGLVPAQMPDARPAIIASSRAGDE 857
Qy 845 PATYLEALGLMAVGLVSNAGLFPSSGRVPLPTYPQORRYMT---DTKADAR-- 898
Db 858 AASALEALGFWGVGVSIVSGVSPSGRRVPLPTYPQORRYMT---DTKADAR-- 917
Qy 899 ----- 898
Db 918 AGGHPLGVEVSSTHAGRLMETTLDKRLPMIGERHAGEVVPFGAGYLEMALSSGAE 977
Qy 899 ----- 904
Db 978 ILDGPIQVTDVLIETLTFAGDTAVPVQVTEERPGRLFOVASREPGRRAPRIHA 1037
Qy 905 ----- 904
Db 1038 RGVLRIRIGVETPARSNLALRLARLHAAPPAIYGALAEMGLQGPALRGIAELMRGEG 1097
Qy 905 ----- 904
Db 1098 EALGRVLPRAAGSATAYQLHPVLLDACVQMIVGAFARDADATMAPEVGVSVLLPQRRP 1157
Qy 905 ----- 904
Db 1158 GELWCHARVVDGQOASSRWGADPELMDGTGAVVAEISRLVVERLASGVRRRADDPLE 1217
Qy 905 ----- 908
Db 1218 LDWEPAALGPKITAGRWLLIGEGGGLGRSLCSALKAGAHVVAAGDDTSTAGMRALLA 1277
Qy 909 ----- 908
Db 1278 NAFDQAPTAIVHLSSLDGGQLGPGIGAQALAPRSVDVADALBESALMRGCDVSL 1337
Qy 909 ----- 908
Db 1338 VQALVGMIDLRNAPRLMLTRGAQAAAAGDVSVQAPLLGLGRTTALAEHAEIARCSVDLP 1397
Qy 909 ----- 952
Db 1398 AEPEGEADALIAELLADAEVVALRGDDLVAARLVHLPDAQREKPEPGRDPPFRLEI 1457
Qy 953 DEPEVLHLVLTERRAPGLGEVEIAVDAAGLSFNDVQLAGVPPDDLPCKPMPPLILG 1012
Db 1458 DEBALQOLVIRATGRAPRGVEIVISVAAGLDSIDLQLALGVAAPNDLPBEELPELVLS 1517
Qy 1013 GECAGRIYAVGEGVGLVGOPIVIALSAGAPATHTTSSAALVLPPOALSAIEAAMPVA 1072

```

```

Db 1518 SECAGRIYAVGEGVGLVGOPIVIALSAGAPATHTTSSAALVLPPOALSAIEAAMPVA 1577
Qy 1073 YLTAAMVLDRIARLOPGERVLIHAATGVGLAAVQMAOHVAGVHATAGTPEKRAYLESL 1132
Db 1578 YLTAAMVLDRIARLOPGERVLIHAATGVGLAAVQMAOHVAGVHATAGTPEKRAYLESL 1637
Qy 1133 GVRVYSDRSRDRFVADVRAVMTGEGVDVVLNSLSEGLIDKSFNILRSHGFVEIGKDCY 1192
Db 1638 GVRVYSDRSRDRFVADVRAVMTGEGVDVVLNSLSEGLIDKSFNILRSHGFVEIGKDCY 1697
Qy 1193 ADNOLGRPLRLNSPESVLDLRCMMLERPARVALLLEELGLIAAGV----- 1239
Db 1698 ADNOLGRPLRLNSPESVLDLRCMMLERPARVALLLEELGLIAAGV----- 1757
Qy 1240 FTPEPIATLPIARVADFRSMAOHGKVLTLTGDEPEVOIRPTHA----- 1286
Db 1758 LTPPEVETPISRAAEFRMAOGHGLKVLTLTDPEVIRAPAESVAVRADGTYLV 1817
Qy 1287 ----- 1289
Db 1818 GGLGLGLRVAGMLAERAGQLVLVGRSAAAEQAAVAALAEHAGARTVAKADVADR 1877
Qy 1290 ----- 1289
Db 1878 QIERVLEVTASGMPLRGVVAAGLYDDGLMOQTPARFRTVMGPKVQALHHTLTREA 1937
Qy 1290 ----- 1289
Db 1938 PLSEFVLYASAGLFGSPGQNTAANAFLDALSHHRAAGLPALSIDKMFTEVGNVA 1997
Qy 1290 ----- 1292
Db 1998 QENRGARQISRMKGTTPDEGLSALRLLEGDRVQGVIPITPRQWVEFPATASRLS 2057
Qy 1293 ----- 1338
Db 2058 RLVTORAVADRATGABDLLEOLASAPASAGILQOVVQVSHVRLTEDKIEVDAPL 2117
Qy 1339 TRIGMDSIMAVELRNRIEASLKLSTTFSTSPNIALLONL-DALATV 1389
Db 2118 SSKGMSLMSLELRNRIEASLKLSTTFSTSPNIALLONL-DALATV 2169

RESULT 3.
US-10-014-717-5
: Sequence 5, Application US/10014717
: Publication No. US20020192778A1
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIIONES
: FILE REFERENCE: 4-30582A
: CURRENT FILING DATE: 2001-11-13
: PRIOR APPLICATION NUMBER: US/10-014, 717
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 7257
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
US-10-014-717-5

Query Match 45.9%; Score 3311.5; DB 14; Length 7257;
Best Local Similarity 41.1%; Pred. No. 1.2e-244; Indels 531; Gaps 25;
Matches 784; Conservative 162; Mismatches 432;
Qy 9 AEDPIAIVGASGRLPGVIDLSGFWTLLEGSRDVTGVRPAERWDAAMPDPDAPGKT 68

```

```

Db 3021 ASDEIATVGAACRPGVEDELESYMOJLABGVVSAEVPADRMADMDDDPEIIPRT 3080
Qy 69 PYTRASFLSDVACPDASFFGISPREALRMDPAHLLILEVCWEALENAIAPSAIVGTETG 128
Db 3081 YTKAFAPRLDLORLDATEFRISPREAMSIDPOORLLEVSWEALESAGIAPDRLDSPTG 3140
Qy 129 VFIGGSEBEAALQOATASAEIDAHGIGTWPVGAGISVALRGCVAVDRAVSS 188
Db 3141 VFGVGPBYTYQRURNGFTDGAAGLYGGGNMLSVTAGRLSFLJHGFTLMDTRCSS 3200
Qy 189 LVAHACQSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARDGRCAFSAEADGFR 248
Db 3201 LVALLACQSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARDGRCAFSAEADGFR 3260
Qy 249 GEGCAVVLKRLSGARADGRILAVIRGSAIINHDASSGLTVPNGSSOEIVLKLADAG 308
Db 3261 GEGCAVVLKRLSGARADGRILAVIRGSAIINHDASSGLTVPNGSSOEIVLKLADAG 3320
Qy 309 CAASVGVYEAHGTGTGDIPIEIOALNAVVGIGRDVATPILIGSVKTMIGHPEVASGIT 368
Db 3321 VSPVDVFECHGTGTGDIPIEIOALNAVVGIGRDVATPILIGSVKTMIGHPEVASGIT 3380
Qy 369 GLLKVLSTLOHGOIPAHLLAQAALNPRISWGLRLTVTRATPMPDMTPRRAVGSFGMS 428
Db 3381 SLLKVLNLRHQIPAHLLAQAALNPRISWGLRLTVTRATPMPDMTPRRAVGSFGMS 3440
Qy 429 GTNAHVLEBEAPATCTPPAPERPAELVLSAKTASALDAQAARLDHLETPSCLODV 488
Db 3441 GTNAHVLEBEAPATCTPPAPERPAELVLSAKTASALDAQAARLDHLETPSCLODV 3500
Qy 489 AFSLLTTSAMEHRLAVATSRERGLRALDAAAOQOTSPGAVRSTIADSSRGTLAFLTGO 548
Db 3501 AFSLLTTSAMEHRLAVATSRERGLRALDAAAOQOTSPGAVRSTIADSSRGTLAFLTGO 3560
Qy 549 GAOTLGMGRGLYDVMSAFREAFDLCVRLFNCLEDRPLREVMMAEPASVDAALLDOTAFQ 608
Db 3561 GAOTLGMGRGLYDVMSAFREAFDLCVRLFNCLEDRPLREVMMAEPASVDAALLDOTAFQ 3620
Qy 609 PALFFETVALAALMSKGVPELVAGHSGIGELVAAQVAGFSELEDAVFLVAARGLMQAL 668
Db 3621 PALFFETVALAALMSKGVPELVAGHSGIGELVAAQVAGFSELEDAVFLVAARGLMQAL 3680
Qy 669 PAGGMVSIIEAPEAVAAVAPPAASVSIAAVNADDOVITAGAGPVAHIAAAMARGR 728
Db 3681 PAGGMVSIIEAPEAVAAVAPPAASVSIAAVNADDOVITAGAGPVAHIAAAMARGR 3740
Qy 729 TKALVSHAFFHSPLMAFMLAFGRVAESVYRPSIVLVSNLSGRACDDEVSSPGYWRH 788
Db 3741 TKALVSHAFFHSPLMAFMLAFGRVAESVYRPSIVLVSNLSGRACDDEVSSPGYWRH 3800
Qy 789 AREVVRPADGVKALAAAGAGTFVEVGPSTLGLVAPCMPOARPLLASSRGRDEPATV 848
Db 3801 AREVVRPADGVKALAAAGAGTFVEVGPSTLGLVAPCMPOARPLLASSRGRDEPATV 3860
Qy 849 LEALGLMAVGLVSMAGLFPSSGRRVPLPTVPMORERYIDTAKDA----- 896
Db 3861 LEALGLMAVGLVSMAGLFPSSGRRVPLPTVPMORERYIDTAKDA----- 3920
Qy 897 ----- 896
Db 3921 GVGLCMGAVLHVLSIGRHPFLGDHLVFGKVVPQAFAHVAVLSTIAEKWPERAIEL 3980
Qy 897 ----- 902
Db 3981 TGVEFLKAIAMEPDEVEHIAVLTPBAAGDGLFEIATIAAETERRMTTHARGVOPD 4040
Qy 903 -ARGA----- 906
Db 4041 GARGGALPRLEVEDRAIOPLFAGFLDRLISAIRIGKGLMWLODGRVDEASLTLVPT 4100
Qy 907 ---GHD-----EVEEGA-----VR-GGD 921

```

```

Db 4101 YPNADVAFLPILLDNGFAVSLSTRSEPEDDGPPLPFAVERVWRMARFVGRCCGV 4160
Qy 922 RSA-----RLDHP----- 932
Db 4161 PRSQAFVSSFLVDETEGEVVAEVEGFCRPAEPREVLROESGASTAALYRLDWEABLP 4220
Qy 933 -----ESGR 936
Db 4221 DAPARIESVWVAABGSEVAAALATRLNRCVLAEPKGLEAALAGVSPAGVICMEGA 4280
Qy 937 REKVEAADR----- 946
Db 4281 HEEBAPAAQFVATGELSVQALRDRAVRLMWWTGAVAVEAGERQVATAFVWGIGRTVM 4340
Qy 947 ----- 946
Db 4341 QERPELCTLVLEBEADAARSADVLRELGRADDETOVAFRSGRRVRLVKATTPEGL 4400
Qy 947 -----PERLEIDEPVLDHVLVLRTERBARGLGEVEIAVDAAGLSPNDVOLAGMVPDD 1000
Db 4401 LVPDABESRLBAGQKTIDQLRLAPQORRABGPEVEIKTASGLNFTTILAVLGM----- 4456
Qy 1001 LPRKPNPLLIGBECAGRIYAVGEGVNGLVVGOPIYALASAGAFTHVTSALVLRPQA 1060
Db 4457 YPGDAGP---MGDCAGVATAGGVVRHVAAGDAMTL---CTLHRFVTVDARLVVRQDAG 4511
Qy 1061 LSAIEAAMPVAYILTAWALRIRARLOPGEVVIHAATGVGLAAVQAOHVGAHVATA 1120
Db 4512 LTPQAOAAVPAFATAMALHDNLNLRGEVVLHAAAGVGMMAVQIARVIGAEVETA 4571
Qy 1121 GTPERKAYLESIGV--RVSDSRSDRFVADVAVRAMTGGEGVUVVNSGELIDKSFNLLR 1178
Db 4572 -SPSKMAAVQMGVPRTHIASRTLEFPAETRYOTGGRGVUVVNLALAGEVVDASLISLS 4630
Qy 1179 SHGRFVELKRDCCYADNQ--LGLRPFNLISFSLVDLRGMLRBPARRVALLBELGLIA 1236
Db 4631 TGRFLTEMGKTD-IRDRAVAHAAPGVYRVFIDLEL-----ABDRREILERVEGFA 4683
Qy 1237 AGVFTPEPIATLPARVADARMSAQAHLKLVLTGLDEPEVQIRIP-----HAGAP- 1290
Db 4684 AGHRLALPVHAFATTKBAARFMAQAHOQKVL--LPAPAAPLAPGTVLLTGGLAL 4742
Qy 1291 -----STGDRDLDRLASAPARAAALBAFLRTQVSOVLRTPEIKV 1333
Db 4743 GLHVARMLAQQGVHMLVTGRRG--LDTGPAKAVAEIETALGARTIASDVADNNALEAV 4801
Qy 1334 AEAL-----FTRLGMSIAMEVLEARNRISLKLKUST 1365
Db 4802 LQATPAEWPLQGVIIAAGALDDGVLEQTTDRFSRVLAPKVTGAMNLIHETLAGNDLAFV 4861
Qy 1366 TELSTS-----PNTALLAQNLLDALATLSLERVAAEVLRAQVOND 1406
Db 4862 LFSSMSGILSAGOSNTA-AANTFLDUALAHRAEGLAQSLAMGPMSD 4909

```

RESULT 4  
US-10-156-761-10436

Sequence 10436, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 10436  
 ; LENGTH: 6146  
 ; TYPE: PRF  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-10436

Query Match 32.1%; Score 2311.5; DB 15; Length 6146;  
 Best Local Similarity 32.3%; Pred. No. 9.5e-168;  
 Matches 633; Conservative 178; Mismatches 519; Indels 631; Gaps 37;

QY 5 PIERAADPIAIVAGACRLPGVIDLSGFMTLEGSRDIVGRVPAER-WDAAMFDDPD 63  
 DB PTAEAADDPVIVAMACRFPEATSPKLMNDLIAAGDGGIDFPADRGWVA-----D 4036  
 QY 64 APGKTPVTRASFSDVACFDASFFGISPREALRMDPAHLLLEVCWALENAALAPBALV 123  
 DB AARS---QTGGFLPDVAEPFAGFPGISPREALAMDPOORLLLETSMWELERAGVDALKLR 4093  
 QY 124 GTENVFFIGPSEYEALPQATASAEIDAHGIGTMPVSGAGISVALIGRGCVAVD 183  
 DB GSRTGVFVAGSHDGTLLTSLSGGDYALTGAVG---SVLSKRIAYVLGLEGPALTVD 4150  
 QY 184 AYSSSLVAHLACQSLRSGCSTALAGVSLMSPSTLWLSKTRALARDGRCAFAEA 243  
 DB ACSSSLVALHLAALAGLGEEDVALAGVAMATPDAFPAAGCGLAPEGRCACAFADGA 4210  
 QY 244 DGRGEGCCAVVILKRLSGARADGRLIIVIRGSAIINHDCASGLTVPNSSQEI VLKRA 303  
 DB DGGWEGGVVLVTLRLSEARRRGHDVIAVVRGSAVNSDGSNGLTLPNGPQQRVTRQA 4270  
 QY 304 LADAGCAASSGVYEAGHTGTLGDPLEIOLANVYGLGRDVAPLLIGSKYKMLGHPEY 363  
 DB LASHGLSADVDMEAGHTGSLGDPLEIOLALTYQODRPAERPLWGSYKSIYGTQA 4330  
 QY 364 ASGITGLKTVLSTOHQIPAHLLHAQALNPRISWGDRLITVTRARTPMDMTERRAGVS 423  
 DB AAGVAGVYKSLVALRNLGRLPTLTHVDEPSREVDSAGAVELLTEGRMPEIDGRRRAGVS 4390  
 QY 424 SFGMSGTNAHVLEAPAA--TCTPPAPERPAELIVTSARTASLIDQAARLRLHETYP 481  
 DB AFGISGTNAHVILEQAPSVEEPVTPAPADDGLVPMVVSARSEELRAQARLADHVAA--T 4449  
 QY 482 SQCGDVAFSLATRSMEHRLVAATSREGLRALDAAGOCSPRANVSIAOSSGKL 541  
 DB DLTVADVGLSLAGTRAGLEHRAVLTGREDEFLVQLAALAGASAGAVRGVA--GEGT 4507  
 QY 542 AFLFTGGAGTGLGWRGLYDVWASAFREAFD-LCYRLFNOELDRPLREVMMAEPASVDAAL 600  
 DB AFLFTGGAGTGLGWRGLYDVWASAFREAFD-LCYRLFNOELDRPLREVMMAEPASVDAAL 600  
 QY 4508 AFLFTGGAGTGLGWRGLYDVWASAFREAFD-LCYRLFNOELDRPLREVMMAEPASVDAAL 600  
 DB AFLFTGGAGTGLGWRGLYDVWASAFREAFD-LCYRLFNOELDRPLREVMMAEPASVDAAL 600  
 QY 601 LDQTAFTOPALFTFEVALAALMRSGVPELVASHGISGELVAACVAGVFSLEDAVFLVAA 660  
 DB LDRVTFVQAGIFALEVALFELVSSWGVADVLGHSIGELAAAYVAGWVSLADACRIVAA 4621  
 QY 4562 LDRVTFVQAGIFALEVALFELVSSWGVADVLGHSIGELAAAYVAGWVSLADACRIVAA 4621  
 DB LDRVTFVQAGIFALEVALFELVSSWGVADVLGHSIGELAAAYVAGWVSLADACRIVAA 4621  
 QY 661 RGRMLQALPFGGAMVSTEARPADVAAVAAPHAASVSTAAYVAPROVITAGAGOVHATA 720  
 DB RGRMLQALPFGGAMVSTEARPADVAAVAAPHAASVSTAAYVAPROVITAGAGOVHATA 720  
 QY 4622 RGRMLQALPFGGAMVSTEARPADVAAVAAPHAASVSTAAYVAPROVITAGAGOVHATA 720  
 DB RGRMLQALPFGGAMVSTEARPADVAAVAAPHAASVSTAAYVAPROVITAGAGOVHATA 720  
 QY 721 AMARGARTALAHVSHAFSPILMAPMLEAFGRVAVESVYRPSIVLNSGKACTOEV 780  
 DB AMARGARTALAHVSHAFSPILMAPMLEAFGRVAVESVYRPSIVLNSGKACTOEV 780  
 QY 4678 TPAEQGRITQOLAVSHAFSHARMEMPLAEFAETLAAVEFRSPRPVPSVNTVGTGAAGAEFT 4737  
 DB TPAEQGRITQOLAVSHAFSHARMEMPLAEFAETLAAVEFRSPRPVPSVNTVGTGAAGAEFT 4737  
 QY 781 SPGVWVHABEVRVAFADGVKALHAAGATVEVGPCKTL-----LGLVPAKMDA 830  
 DB TPAEQGRITQOLAVSHAFSHARMEMPLAEFAETLAAVEFRSPRPVPSVNTVGTGAAGAEFT 4737  
 QY 4738 TPAEQGRITQOLAVSHAFSHARMEMPLAEFAETLAAVEFRSPRPVPSVNTVGTGAAGAEFT 4737  
 DB TPAEQGRITQOLAVSHAFSHARMEMPLAEFAETLAAVEFRSPRPVPSVNTVGTGAAGAEFT 4737  
 QY 831 RPLALLASBRGRDPATVLEALGLMAVGLVSNAGLFPSSGRVPLPTVPMOSERWID 890  
 DB RPLALLASBRGRDPATVLEALGLMAVGLVSNAGLFPSSGRVPLPTVPMOSERWID 890  
 QY 4798 HPE-----RPPDATTLLHALATVATVGAADVMTTPP-ACARTTELPTVAFQRRRW-P 4847

QY 891 TKADDAAGRRAPG---AGH----- 908  
 DB TKADDAAGRRAPG---AGH----- 908  
 QY 4848 RPAASAGADLAAAGLAEGHPLTAMLPSPGEGVLCTGRSLATHPWLGHTVLTCTVL 4907  
 DB RPAASAGADLAAAGLAEGHPLTAMLPSPGEGVLCTGRSLATHPWLGHTVLTCTVL 4907  
 QY 909 ----- 908  
 DB ----- 908  
 QY 4908 VPGTAFLVDLVGVGDRVCGGIRELTPLALAEBAVRLRVVLGAPDDTCRPAVYA 4967  
 DB VPGTAFLVDLVGVGDRVCGGIRELTPLALAEBAVRLRVVLGAPDDTCRPAVYA 4967  
 QY 909 ----- 908  
 DB ----- 908  
 QY 4968 QPDTDEGTWTHAEGGLAPADSDTPRAPRPPALDAMPVPGAEPVPMGFEGLADAG 5027  
 DB QPDTDEGTWTHAEGGLAPADSDTPRAPRPPALDAMPVPGAEPVPMGFEGLADAG 5027  
 QY 909 -----DEV-----EEGA----- 916  
 DB -----DEV-----EEGA----- 916  
 QY 5028 FAYGPLEFGRRAARRRDEVEFAVSLPADSSGAGFVHPALDLAALHAGPAQAASDEPG 5087  
 DB FAYGPLEFGRRAARRRDEVEFAVSLPADSSGAGFVHPALDLAALHAGPAQAASDEPG 5087  
 QY 917 -----VRGD----- 921  
 DB -----VRGD----- 921  
 QY 5088 SARLPSPSGEVRVHAGADLLRVLRVRAADGVTTLDAADAGRPVVSIGSLVRLSPDR 5147  
 DB SARLPSPSGEVRVHAGADLLRVLRVRAADGVTTLDAADAGRPVVSIGSLVRLSPDR 5147  
 QY 922 ----- 921  
 DB ----- 921  
 QY 5148 LRAQTAQPDALPATKPIPLDVTGDTGDTARDTVECVLGEPSAGAMRSHPDAAVAAAL 5207  
 DB LRAQTAQPDALPATKPIPLDVTGDTGDTARDTVECVLGEPSAGAMRSHPDAAVAAAL 5207  
 QY 922 -----RSAR----- 926  
 DB -----RSAR----- 926  
 QY 5208 ADGKETPGVPLARCPITTGDPAAARSAEWTLDLQWLAADRLTDSHLVIGTRHAPV 5267  
 DB ADGKETPGVPLARCPITTGDPAAARSAEWTLDLQWLAADRLTDSHLVIGTRHAPV 5267  
 QY 927 -----LDHPPESGREKEYBA----- 943  
 DB -----LDHPPESGREKEYBA----- 943  
 QY 5268 ATPYAAPVDATDADAAFPVDAGHEAPVDTAAALAQSTLAGLVSAQOTENPARITLVDPDT 5337  
 DB ATPYAAPVDATDADAAFPVDAGHEAPVDTAAALAQSTLAGLVSAQOTENPARITLVDPDT 5337  
 QY 944 -----GDRPRLIDEPPGL 959  
 DB -----GDRPRLIDEPPGL 959  
 QY 5328 TAEDPAHLTAAYVRLGEBEVAVRQGTLYARLTPETGRALTVPPEGPWMLDSTGRSLD 5387  
 DB TAEDPAHLTAAYVRLGEBEVAVRQGTLYARLTPETGRALTVPPEGPWMLDSTGRSLD 5387  
 QY 960 HLIVRTERRAPGL--GEVEIADVAGLSFNDVQLAGVPPDDLPGKPNPPLLGGECAG 1017  
 DB HLIVRTERRAPGL--GEVEIADVAGLSFNDVQLAGVPPDDLPGKPNPPLLGGECAG 1017  
 QY 5388 HLIVPCPDAAALAEQTVIAVHAAGVNPFRDILALDM---YGRAD---LGTGECAG 5439  
 DB HLIVPCPDAAALAEQTVIAVHAAGVNPFRDILALDM---YGRAD---LGTGECAG 5439  
 QY 1018 RIYAVEGNGVLYVQGVYVLSAGAPATHVTTSAALVLPFQALSALEAAMPAVYITAM 1077  
 DB RIYAVEGNGVLYVQGVYVLSAGAPATHVTTSAALVLPFQALSALEAAMPAVYITAM 1077  
 QY 5440 VLEETGGVYGLAPGRWKMVMVAGFPPTAVADRVLARIPDMSFETAATPVAFLTAY 5499  
 DB VLEETGGVYGLAPGRWKMVMVAGFPPTAVADRVLARIPDMSFETAATPVAFLTAY 5499  
 QY 1078 YALDRJARLOPGRVYIHAATGGVGLAAYOVAGVEVATGTPREKAYLES-LGVRY 1136  
 DB YALDRJARLOPGRVYIHAATGGVGLAAYOVAGVEVATGTPREKAYLES-LGVRY 1136  
 QY 5500 YGLVLDLGLGAGSGLVHAAAGVGMVAQVLAHNLGHEVGTASBPXKSTLDTGLDRAH 5559  
 DB YGLVLDLGLGAGSGLVHAAAGVGMVAQVLAHNLGHEVGTASBPXKSTLDTGLDRAH 5559  
 QY 1137 VDSRSDRFVADYRAMTGEVGVVINSLSGELIDKSFNLLRSHGRFVELGKDCYADNQ 1196  
 DB VDSRSDRFVADYRAMTGEVGVVINSLSGELIDKSFNLLRSHGRFVELGKDCYADNQ 1196  
 QY 5560 IGSSRTLDPADLVRETTGAGGVVINSLSGELIDKSFNLLRSHGRFVELGKDCYADNQ 1196  
 DB IGSSRTLDPADLVRETTGAGGVVINSLSGELIDKSFNLLRSHGRFVELGKDCYADNQ 1196  
 QY 1197 LGL-RPFLRNLSGLVDLRGMLEPARVALLLEELGLIAAGVFTPPPLATLPIARVAD 1255  
 DB LGL-RPFLRNLSGLVDLRGMLEPARVALLLEELGLIAAGVFTPPPLATLPIARVAD 1255  
 QY 5620 VAADHPEVRVRSFPLDG-----AGPRLIAELIAHVELFESGALTPLPVTTDIRDATD 5673  
 DB VAADHPEVRVRSFPLDG-----AGPRLIAELIAHVELFESGALTPLPVTTDIRDATD 5673  
 QY 1256 AFRSMQAOHLGLVYTL----- 1273  
 DB AFRSMQAOHLGLVYTL----- 1273  
 QY 5674 AFRVLSQATLTGKAAVLTVPAGSVPFPAEGFTVLITGGTGLTGLTLARHLVTEHGRHYTLA 5733  
 DB AFRVLSQATLTGKAAVLTVPAGSVPFPAEGFTVLITGGTGLTGLTLARHLVTEHGRHYTLA 5733  
 QY 1274 -----GDPEV-----QIRIPTHAGSPSTGDRDLDD-----SLASAPARA 1310  
 DB -----GDPEV-----QIRIPTHAGSPSTGDRDLDD-----SLASAPARA 1310  
 QY 5734 GRGTDTPFVRQRLARDALEMAEVRVA-CDADDEDAVRLRLBELTANHRLAGVHAAV 5792  
 DB GRGTDTPFVRQRLARDALEMAEVRVA-CDADDEDAVRLRLBELTANHRLAGVHAAV 5792  
 QY 1311 ---AALEAPLRTVOVSLRTPPEIKVGAELFTLGDNDLSMAVELRNRLEASLKXKSTTF 1367  
 DB ---AALEAPLRTVOVSLRTPPEIKVGAELFTLGDNDLSMAVELRNRLEASLKXKSTTF 1367  
 QY 5793 TDDGVVSLDRRLRLSLVLAH-PKVR-GAMNL-----HRLTANLAPRMVLF 5835  
 DB TDDGVVSLDRRLRLSLVLAH-PKVR-GAMNL-----HRLTANLAPRMVLF 5835  
 QY 1368 LSTSPNIALAON-----LDALATALSLEVAENLRAGV 1403  
 DB LSTSPNIALAON-----LDALATALSLEVAENLRAGV 1403  
 QY 5836 SSASATTTGAAGGQVYAAANAFUDALAEHRRARGLPSSLAWGL 5878

RESULT 5  
US-10-156-761-7962  
Sequence 7962, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 7962  
LENGTH: 6145  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-7962

Query Match 31.2%; Score 2249; DB 15; Length 6145;  
Best Local Similarity 37.2%; Pred. No. 6.2e-163;  
Matches 611; Conservative 183; Mismatches 545; Indels 304; Gaps 50;

QY 1 VADPRIEAABD-PAIVAGSCRLPGVYIDISGFVTLLEGSDTQGRVPAER-MDAAMWF 58  
DB 1531 VVEQOHAADDEPAIVAMSGHLPFGVDSPEALMDLVTSQGDALISEPNSRGMVDVLY 1590  
QY 59 DDDPAKTEVTRASFSDVACFDASFFGISPREALMDPAHRLLEVCWEALENAAIA 118  
DB 1591 DDDPRPKTYABDQGLYATDFDAGFPFGISPREALMDPQRLMLETSWEAFERAGIA 1650  
QY 119 PSALVGTGTYGFIGGSEYEALPQATASAEIDHAGLGITMPSYAGARISYALGRPC 178  
DB 1651 PAQGRSGTYGVFGMAVYQGYGADVARTPEG--VEGHRLVGASVSVSGRVATYFLEBPA 1708  
QY 179 VAVDTAASSLVAVHACQSLRSRGCSTALAGVSLMSPSTLWVLSKTRALARDGRCKA 238  
DB 1709 VTITACSSLVALLHMQSLRNGECTVALLAGGVTVMAGPSYFVEFSQKGLSPDGKRA 1768  
QY 239 FSAEDGFRGEGCAVVVLKRLSGARADDRILAVIRGSAINHDGASGLTVPNGSSQEI 298  
DB 1769 FGADADGTGMAEGVGMILVERLSDBRRNGHVLAVVRSVAVNQDASNGLTAPNGPAQR 1828  
QY 299 VLKRALDAGCAASSVGVYEAHGTGTTIGDPIEIOALNAVGLGADVATPLLIGSVKTL 358  
DB 1829 VIHOSLSAGLAPRDVAVEAHGTGTRLGDPLEAALATYGGCGDGRPTMLGSLKSN 1888  
QY 359 GHPEVASGITGLKLVLSLQGOIPAHLLAOLNRLISM--GDRJLVTRARTPMPDMNT 416  
DB 1889 GHTQAAAGVAVIKVMAMRGILPRLTHADEPTPHVMSAGDVLL-ITEA-VDMPESDR 1946  
QY 417 PRRAVSSFGMSGTNAHVLEBPAPATCTPPAPERPAEL-LVLSARTASALDAQARLD 475  
DB 1947 PRRAISFGVSGTNAHTIIEQAPATAELPTPDGALVPMVLSKGSPALRAQAARLD 2006  
QY 476 HLETPSCGLDVAFSLATTSAMEHRLAVALAATSREGIRALDAQAQOSTSPGAVRSIAD 535  
DB 2007 HLDANDGTASDIGSLA-SRETFEDRAVLLAOGSDKRLTALAQAQOPANVTLGRA- 2064  
QY 536 SSRGLATFLTGQAGQTLGMRGLVDVMSAFREAPD-LCVRLFENGLDRPLREVMAEPA 594  
DB 2065 RSEKGVGLFSGGQSQIRIGMRELYEAFPVADADIVECARL-----EAPV 2110  
QY 595 SVDAALDQATFTQPALTFEYALALALWRSWGVPELVAHSGISGELVACVAGVFSLEDA 654

DB 2111 DVDAETLHRTGCAQPALFAVEVALFRLLEGWGVDPDYVAGHSGEIAAAHVAGVLSLDA 2170  
QY 655 VEIVAAARGRLMOLPAGGAWSTIEAPADVAAAAPHA-SVSTAAVNAPOQVVIAGAQ 713  
DB 2171 VRLVSARPAALMOLPVGAMVAVQATEE---VLPHLTBEVGTALINGPSMWSLGAGE 2226  
QY 714 PVHIAAAMARGARTALHVSASHPLAPMLAEFAGVAESVYRRPSVLVSNLSGK 773  
DB 2227 AVTIAAEVFGQKGRSLKXSHAFHSPLMDPMLIEFAEVVRGLTFHQPOVPVVSNTLGR 2286  
QY 774 ACTDEVSSPGYVTHAEVVRPADGVKALHAAGTVEVGPSTLLGLVPACMPDA--R 831  
DB 2287 LA--EPYSPETWYHVEAVRFADGIELTGELTGTVEIGPGSVLSGMVQDCVDEAVTV 2344  
QY 832 PALLASRAGDEPATYLEALGMAVGVSVAGLPPSGGRVPLPTYQRRRWYMDT 891  
DB 2345 PVL---RGDRPEQAAVVTALAEHLHGVSPDWALPP-GAHRVTLPTVAFQRRFMDG 2399  
QY 892 KADDAARG--DRAPGAGHDEVE---EGGAVRGDRRSARLDH--PPESGREKYE-- 941  
DB 2400 EGDLAAGTPTDADSGFMSVEREDAESLAATLGVSADASLGALLPRLSMWRQRRQS 2459  
QY 942 ---AAGDRPF---RLTDEPGVLDHLVLYTER 968  
DB 2460 VVDGMRVYVWKPLGALPQPGASGTWLLVVAESEMWTASVRTALDERGL--ELVTLLVA-- 2515  
QY 969 RAPGLGEVETAVDAGLSFNDVQALGAVPDPLDGKRNPLLLGEGCAGRIVAVGEVNG 1028  
DB 2516 -GPDTDGTLLRELAGV--GPVGTVLSLAEDESASGDPGLSGLAH--TLCLVQALGD 2570  
QY 1029 LVWQPVYALSAGAFATHVT-----TSAALVLR-----POLSAIE 1065  
DB 2571 VGVDPPLMCATRGAVNAIGRSRVDRLPLOSQWCGRAALEHPRMGGLDLDPDLART 2630  
QY 1066 AAAMPVAVLTAVALDRIA-----RL-----OPGERVLIHAATGVGL 1103  
DB 2631 AAR--VAAVIGQTEDDQVAVASGVLRRLVRAARATGPAQWSPRGVTLTGCTGALGG 2688  
QY 1104 AAVQWAGHVGAEVH-----ATAGTPKRAVLESIGVR-- 1135  
DB 2689 HVARMLAGAGAE-HLVITSRRGPAAPGAELVVAELIEGAATVAVCDADRDLRALLLA 2747  
QY 1136 ---YVSDRSRDRFVADVRA----- 1151  
DB 2748 QHPVNAVHAAVGDHMIEDSDPAGFAGSVAAGAATHDELLAGOELDAFVMPSSGA 2807  
QY 1152 --WTG--GEGVDVVLNSLSGELIDKSFNLRSRGR-----FVELGRDCTVADNQLGL 1199  
DB 2808 GIMWGAGQAGVSAANAVLIDLALAEHR---RAHGRTALAVSWGVAEGMAG-VGDGEBML 2862  
QY 1200 RPLRLNLSFSLVDRGMMLERPARVRLLEEL--GLIAGV-----FTPPPIATLPI 1250  
DB 2863 R---RRGLPRPKPALAISALQOALADGETALTADVDMERFLGPTVTRPS 2910  
QY 1251 ARVDAFRSMAQOHLKLVLTLDPEVOIRIPHAAGSPSTGDRDLIDRLASAPARA 1310  
DB 2911 ALFSDV-----PEAGRALAPAGAGAEAGNA-LAAKLTGLPAABQD 2950  
QY 1311 AALEAFRLTOVSQVL---RPEIRVGAALFTRLGMSLMAVEIRIEMSLKSLSTTF 1367  
DB 2951 RTLVLDVLRTHAAVLTGHDGTAAPVPG--RAFNDJGFSLSLAVELRNKLTLDTGKLP TTL 3008  
QY 1368 LSTSPNIALLAQ---NLLODLATALLSLERYAAE-----NLBAGVON-----D 1406  
DB 3009 VEDYFNATLARFLRAELLSRGARQORLAAVGDDEPIALVMSCHLPQGVSGSPEDLMQ 3068  
QY 1407 FVSSGA-----DQDMEIATL 1421  
DB 3069 LVTSQGVISGFPEDRGMVEAL 3091

RESULT 6  
US-10-014-717-6

```

; Sequence 6, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Geerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-10-014-717-6

Query Match      31.1%; Score 2245.5; DB 14; Length 3798;
Best Local Similarity 36.2%; Pred. No. 5.6e-163;
Matches 605; Conservative 191; Mismatches 495; Indels 379; Gaps 43;

QY 9 AADDPATVAGASRLPGGVLDLGSFWTLLEGSRTDVGVPABRMDAAAFDPDPDAGKT 68
DB 1519 AADDDATVAGAACRFPGDEGLETYWRHLAEQNVSTVEPADWRKRAADWDPPPEVGGRT 1578
QY 69 PYTRASFSDVACDFDSFPGISPREALRMDPAHRLLEVCWEALENNAIPASLVGTETG 128
DB 1579 YVAKGAFLRVRLDLDAFAFISPREMSLDPOQRLLLEVSMEALERRGQPMALRBSATG 1638
QY 129 VFIGIGBSEYEAALPQATASAEIDAHGGLGTMBPVGAGRISVALGRPCVAADTVAYSS 188
DB 1639 VFGMGISEHAERVOGLDDDAAL-LYGTGNLTVSAAGRSLFPGILGPTMTVDIACSSS 1697
QY 189 LVAVVHLACOSLRSGECTALAGVSLMSPSTLVMTSKTALARDGCKAFSAEADQFGR 248
DB 1698 LVVHLACOSLRSGECQALAGSSVLSRSPFVAASRMRLSPDGCCKTFSAAADGPAR 1757
QY 249 GEGCAVAVLRLSGARADGDRILAVIRGSAINHDSAGSGGLTVPNGSSQEIIVLRKALADAG 308
DB 1758 AEGCAVAVLRLKLDADQGRDPIILAVVSTAINHDGSGGLTVPSGPAQALRLQALAAQ 1817
QY 309 CAASVGVYEAHGTGTTLDGPIEIQALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
DB 1818 VADAEVDVECHGTGTFALGPPIEQALGAVYGRGRPAERPLMLGAVKANLGHLEAAAGLA 1877
QY 369 GILKVVLSLOHGOIPALHLAQAALNPRIKMDLTLTVTRATPMDMTTPRACVSSSGMS 428
DB 1878 GVLKVVLSLHEHGOIPQPELDELNPHIPMAELPVAVVRVAVPWRGARPRAGSAGLS 1937
QY 429 GTNAHVLEBAPATCTPPAPERPAELVLSARTASALDAQAARLRDHLTYPSQCLGDV 488
DB 1938 GTNAHVLEBAPAVEPVAAPERPAELFVLSAKSAAALDAQAARLRDHLTKHVBELGIDV 1997
QY 489 AFSLATTRSAMEHRLAVAAATRSREGLRAALDAQAQOTSPGAANSIAD--SSRGKLAFLFTG 547
DB 1998 AFSLATTRSAMEHRLAVAAATRSREGLRAALDAQAQOTSPGAANSIAD--SSRGKLAFLFTG 2057
QY 548 OGAGOTLGMGRGLYDVMSAFREAPDLCYRLFNQELDRDLREVMAEPAVSVDALLDQAF 607
DB 2058 OGAGOTLGMGRGLYDVMSAFREAPDLCYRLFNQELDRDLREVMAEPAVSVDALLDQAF 2114
QY 608 QPALFTFEYVLAALMRSGVPELIVASHISIGELVAACVAGVSLFEDVFLVAARGRLMQA 667
DB 2115 QPALFTFEYVLAALMRSGVPELIVASHISIGELVAACVAGVSLFEDVFLVAARGRLMQA 2174
QY 668 LPAAGAVNSTEADPADVAAVAPPAASVSIANAAPQOVVIAGAGQPVHAIAAAMAARGA 727

```

```

DB 2175 ISGGEMALVEISLEBAEALRGHGRSLVANSNSPRSTVLAGEPAALBEVLAALTRAKV 2234
QY 728 RTVALHVSASFSPPLMAPLEAFGRVAESVSYRPSIVLSNLSGACTDEVSPPGYWR 787
DB 2235 FMRQVAVDVASHSPQDPLREBELIALGAIIRPAAAVPMRSTVTGVIAGPELGASVWAD 2294
QY 788 HAREVVRPADGVKALHAAGTFFVEVGPSTLGLVPACMPDRP--ALLASRAGDEP 845
DB 2295 NLRQVRFAPAAQAALLLEGCPALFIEMSPHIIVPPDEIQTAAEQGAAVGSIRRCQDER 2354
QY 846 ATLEALGLMNVAGVLSWAGLPPSGRRVPLPTYWQERRYVITDKADDAAGDRAPG 905
DB 2355 ATLEALGLTMASGVVSWARLEPPAGGRVPLPTYWQERRYVITDKADDAAGDRAPG 2406
QY 906 AGHDEVEEGAVAGGDRSARLDHP-----EDSVHGSK--- 930
DB 2407 -----PSLRRLRLRGATDHPILGAPLVARSAPGALHMQALSDERLSYLSHRVH 2457
QY 931 -----PPSGRRREKVEAGDRPPRLLEIDEPG---VIDHLVL-----RVTE--- 967
DB 2458 GEAVLPSAAVVEWALAAQ-----VDLYGTATLVLEQLALERALAVPSEGRIVQVALS 2510
QY 968 RRARGLGEVBIAN-DAAGLSFNDVOALGMV-----PDDLPGK 1004
DB 2511 EEGPGASQVSSREBAGRSW--VRHATGVCSGSSAVGALKEAPWEIQRCPVLSSE 2568
QY 1005 PNPPL-----LG-GECAGRIVAVEGVNGLVVGQPVIALSAGAFAT 1045
DB 2569 ALYPLNHAIDYPCFGQVEQVWLTGTGVLGRVLRPGD-----MASSGAYRI 2617
QY 1046 H-----VTTSAALV-----LPRP-----QALSAI----- 1064
DB 2618 HPALDLACQVLTALITTPRESIERRLRLDHLDPDRSRAPVNQAVSDTWLMDALDGC 2677
QY 1065 --EAAAMPV-----ALTATYALDRIRAR-----LQGERVLIHAATGVGLA 1104
DB 2678 RRGASVAVPDLVLSGFHAKWEVMERLAQAYITGLTIWNVFCAAGSRHTIDELLVQIS 2737
QY 1105 AV-----QWAOH-----VGAEVHATAG-----TEPKAYLESIGRVYSDSRDFVA 1147
DB 2738 VYRKVIKXKMEHLVAIGLVGDGEHFVSSQPLPEPDLAAVLEAG-----RVFA 2787
QY 1148 DVRA--WT--GEGVDVVL-----NSLSGELI 1170
DB 2788 DLPLVLEWCKFAGERLADVLTKTLALBELPFGSGFDMAERIVRDSPIARYSNGIYRGV 2847
QY 1171 DKSFNLRSHGR--VELG-----KRDCT-----ADNQLGLR 1200
DB 2848 ESARVAVASGMSFSLIEIGATGATTAAYLVPLLPRTEYHFTDVSPLFLARAEQRRDY 2907
QY 1201 PPLRNLFSFLVD-----LRGMLEPRPARVRLBELLGLIAGVF-- 1240
DB 2908 PFLK---YGLVDQEPAGOGTAHQRFVIVANVHAITRDIDYATKRLSLSLAPGLLV 2964
QY 1241 ----TPPI-----ATLPIARVADAFRMAQAQHL-----GK 1268
DB 2965 LVEGTGHPIMFIDITTGILBGMQYEDDLRIDHLLPARTCDVLRVGFADAVSLPEDGS 3024
QY 1269 LVITLDPDEVQIRIPFHAGAG--PSTGDRDLDLDAAPAAAPAA-----A 1312
DB 3025 PACILQOHTVLSAPPIAGACDSGES-----ATESPAAARVROEMWADGADVHRMA 3078
QY 1313 LBA--FLRTQVSOVLRTPEIKVGAELPFRIGMSIWAVALRRIEASLTK 1360
DB 3079 LERMTYHRPRGRQVWVHGRLRTGGGA--FKALAGDLLLEFDTCQVVAEVO 3127

```

```

RESULT 7
US-10-156-761-7961
; Sequence 7961, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI

```



APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 CURRENT FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 7961  
 LENGTH: 3352  
 TYPE: PR  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-7961

Query Match 30.5%; Score 2199.5; DB 15; Length 3352;  
 Best Local Similarity 37.2%; Pred. No. 1.6e-159;  
 Matches 611; Conservative 182; Mismatches 529; Indels 321; Gaps 51;

QY 4 RPIERADPPIAIVGASGRLPGVLDLGFITLGGSDTVGRVPAER-NDAAAMPDDP 62  
 DB 27 REVEKDDPPIAIVGACRYPGVRSPPDLMLVAAGDAVDPADGMDVEGIDYDDP 86  
 QY 63 DAPKTPVTRASFSDVACFDASFFGISPREALNMDPAHRLLEVCMEALENAAIAPSL 122  
 DB 87 DAKRTYARRGGFLYEAGEFPDPPFGGISPREAVAMDPOGRILLERTWTTFERAGIDAESV 146  
 QY 123 VGTETGVFIGPSEYEALPQATASAEIDAHGIGTVPYSGAGRIYALGIRPCVAVD 182  
 DB 147 RGSRTGVFGSGYDYMVL--VGVGESDHLGTGNASVMSGRVATFGLGPAVTV 204  
 QY 183 TAYSSIVAVHLAQSLSRSGCSTALAGVSLMSPSLVWLSTKRALARGRCAFAE 242  
 DB 205 TACSSSLVHLAQLNNGECSMALAGVOVMTTPTFAVERSRGRGLAPGRCAFAE 264  
 QY 243 ADGFGREGCAVVVLKRLSGARADGRLIAVIRGSAIHNHDASSGLTVPNGSGOEIVLKR 302  
 DB 265 ADGGMAGVGMLVERISDAVRNGHEVLAVRGSANVDGASNGLTAPNGPAQORVIRQ 324  
 QY 303 ALADGCAASSVGYEAHGTGTTGDPTEIQALNAVGLGSDVATPLLIGSVKTNLHPE 362  
 DB 325 ALASAGLSPAQIDVVEAHGTGTTGDPTEIQALNATYGRERTEDRPLMLGSLKSNIGHSQ 384  
 QY 363 YASGITGLKVVLSLQHQIIPAHLAQALNPRIS--GDLRLTVTRATPMDMNTPRRA 420  
 DB 385 AAGVGGLIKVMAMRHGVLPTLHVDPTPHVDSAGAVRL-LTEA-VEPETHPRRA 442  
 QY 421 GVSSFGMSGTAHVLEBPAPATCTPPAPERPAEL-----VLSARTASALDAQARL 473  
 DB 443 AVSFGVSGTAHTIEQAF-----PAQDEGAGALVDPGLAMVLSASDALRQAARL 497  
 QY 474 RDHL--ETYPQCLADVAFSIATRSAMEHRLVAATSRREGRLAALDAAGQISPGAVR 531  
 DB 498 RPSARGEVRPQ---DIGLSLATTTAAAMRHRAAVVGDREBELSLGLEDLATG--TPSSRV 551  
 QY 532 SIADSSRCKLAFITPGGAOTLGMGRGLYDVMSAFREAFD-LCVTLFNOELDRPLAREVM 590  
 DB 552 LLGRPAGGKTGFLPSGGSORIGMRELYAAYPVFAAAYDEVCAHL----- 597  
 QY 591 AEPASVDALADQAFATFTFEYALALMRSGVPELVAGHSIGELVAACVAVGS 650  
 DB 598 DAPVDVDEFTLRGCTOPALFAVEVALFRLLSEIGVAPDPVGVGSVEIAAAHVAGLS 657  
 QY 651 LEDAVFLVAARGLMQLPAGCAWVSIAPEADVAVAAPHA--SVSIAAVNAPDOVIA 709  
 DB 658 LDDAKLVASABAALMQLPAGCAWVAVOATEE-----VLPHLTDVSIAAVNGPSSVVS 713

QY 710 GAGPVHAIAMAAARGATKALVSHAFSPMLAEAFGRVAESVSRPSIVLVSN 769  
 DB 714 GDETAALAIAGFBQKRTSRKLVSHAFSPMLDPLAEFAVVRGLAEQPLPVSN 773  
 QY 770 LSGRACTDEVSSPGYWRHAREVRFPADGVKALHAPAGATFVEVGPSTLLGLVPACMPD 829  
 DB 774 LTQGPV--EAVTPEYVRHAREVRFPADSVRTLHDLGVKTFIEIGPGVLSGMQGCADD 831  
 QY 830 --AAPALLASSRACRDPATVLEAGLMAVGLVSNAGLFPSSGGRVPLPTVWQREY 867  
 DB 832 IVTVPVV--RADRPERRQAVTALAHHTGVSVDHMSFF-AGARTDLPVAFORBEWY 886  
 QY 888 WIDTKADAA-----RGDRRAPGAGHDEVEGCAVRCGDRSARLDHPPE 933  
 DB 887 WVDAPESAAEADVDVAFETVYERGLQA-----LAETLDGADDTGDLV--FRLS 937  
 QY 934 SGRREKVE--AAGDRPRL-----EIDEPGLDHLVLTERRAPGLGEVEIAVDAAGL 985  
 DB 938 SWRQRKEQSTVDDRVRRESWKRLGELGPAGIGGTWLLAVPAEBEQTAVRATLTARGA 997  
 QY 986 SFNDVQLAGVPPDLFGKRPPLLLGECAGRIVANGEGVGLV--GQPI----- 1036  
 DB 998 TLKT-----LVDDAASHRAG--LAGELAG--IGPVDGLSLVNGDPLVPTLLVQA 1045  
 QY 1037 -----ALSAGAFATH-----VTSAAVLPR-----POL- 1061  
 DB 1046 LGDAGVAPLMCLTSGVAVSGSDAVDARHAQVWGIGRTVLELPRKMGCLDLPESVD 1105  
 QY 1062 --SAIEAAMPVAVLTMAVLAIRIA-----BLQGERVLIHAAT 1098  
 DB 1106 DQAVVRLADVLEQLTGRMEBDQLAVASGVFARLHAAPARAATRMQGRGTVLINGCT 1165  
 QY 1099 GGVGLAQAQAQVHGAEVHA-----TAGTPEKAYLESIGVR--YVSDNSDR--- 1144  
 DB 1166 GALGHAARMLARGAE-HLVTSRRGADAPGAALDELEVIGARVTFACDVAADDAV 1224  
 QY 1145 -----FVADVAMTGGEGVDVVLNSGELIDKS 1173  
 DB 1225 AALLAQHVFAVVAAGVADAGVDTATTPAFAAALAAKYGGA--AHIDELLDQDELDA 1281  
 QY 1174 FNLLSHGRFVELGRDCVADNOLGRLPPLNLSFSLVDRGMWLEPARYRALLBELL- 1232  
 DB 1282 FVLFSISGVWSSGQAAYVAG--NAFLDGLA-----RQRDRGTLATAVS 1325  
 QY 1233 -----GLIAG-----VETPPPIATLPARVADARSMQAQHLKVLTL 1273  
 DB 1326 WGPVADGMAVADGDDEERLRRGLRAMTP-----ASATTSLORALDRDETLTV 1374  
 QY 1274 GD-----PEVO--IRLPTHAGSPSTGDRDLRLASAPAPAR 1309  
 DB 1375 ADVVMSRFIVPTLGRSPPLGDLPEYRALAVDTTPA--AGHSG--ALAESIALSHQDR 1431  
 QY 1310 AALAEFLRQVSGVL-RTPEIKVGAELFTRLGMSLMAVELNRIEASLKLSSTFFL 1368  
 DB 1432 ABALVLDVTRTHAAVVLGHRGAGAVEADRPRDIFGDSLTFAVELNRIKATGTSLPATLI 1491  
 QY 1369 STSPNIALLAQNLLDALATALSIEVAENLRA-----GVON-----D 1496  
 DB 1492 FDHRTAADLAAHLAEL--TGGQHEAVAAATAVAVLDEPIAIIIGMACRYPGGVRSPPDLME 1550  
 QY 1407 FVSSGAD-----ODMEIATL 1421  
 DB 1551 LVASGRDAISRFPNGRMDVEAL 1573

RESULT 8  
 US-10-156-761-7964  
 ; Sequence 7964, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIYUKI  
 APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 CURRENT FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 7964  
 LENGTH: 3564  
 TYPE: PR1  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-7964

Query Match 30.5%; Score 2197.5; DB 15; Length 3564;  
 Best Local Similarity 34.7%; Pred. No. 2.5e-159;  
 Matches 638; Conservative 187; Mismatches 546; Indels 467; Gaps 57;

QY 4 RPIRAAEDPIAIVGASCRIPGVIDLSGFWTLLEGSDTVGRVPAER-WDAAMFDDP 62  
 DB 26 RDVEDADAGEPAIIIGMSCRPGVDSPELWELVAGRDVAEFPADRGMDLDALYDDP 85  
 QY 63 DAPKTPTRASFSDVACPDASPRGSPREALBMDPHRLLETCWCEALENAMAPAL 122  
 DB 86 ESRGTSTVRBGAFFRLDRPDAGLFGISPRELAMPQORLLETSMLEFERRAGIDASL 145  
 QY 123 VGTETGVFIGIGPSEYEAL---PQATASAEIDAHGIGTWPVSGAGRSYALGIRGCV 179  
 DB 146 AGSTGVAGVMNEHFLALQNSPQ---DLEGVLGTGTSGSVASGRVATPGLGEPAY 200  
 QY 180 AVDTAYSSLYAVHILACQSLSGECSTYLAGVSLMSPSTLWLSKTRALARDGRCAAF 239  
 DB 201 TVDPACSSILVAHLAQLVSLRNGECSILVAGGVATMGPAITFVGSFRGLAPGRCRPF 260  
 QY 240 SAADGFERGRCVAVVVKLSGARADGRILAVIRGSAIINHDAQSGSLTVPNGSSORIV 299  
 DB 261 AAGDGTGMBGAGLLVERLSDARRNGHEVALVRSAGVAVNODGASNGLTAPNGSQQRV 320  
 QY 300 LKRALADAGCAASVGVYEAHGTGTTGDPTEIQALNAVYLGIRDVATPLIGSVKTLG 359  
 DB 321 IRQLAAQAQLTTAQVDVAEHGTRGCDPIEBAQLLATYQGRDDKPLWLGSTKSVIG 380  
 QY 360 HPEYASGITGLKVLVLSQHQIIPAHILHAQLNPRISW-GDLRLTVTRARTPMDNTP 417  
 DB 381 HTQAAAGVAGIIKVMAMRHGELPRTLHVDAPSVHVDTAGAVRLLT--ERTPMSASDP 438  
 QY 418 RRACVSSGMSGTNAVVLLEAPAA---TCTPRAPERALL----- 456  
 DB 439 RRAGTSSGSGTNAHTTIEEAPSEAPAOPTGAEGEBEGESAEPVASEAPAPACGP 498  
 QY 457 ---VLASRTASALDAQARL--RDHLETPSQCLGDVAFSLATRSAMEHRLAVALTSRE 511  
 DB 499 VPWVISGTEBALAQAERLLSRDPAEYPL---DVAFSLATRTALEHNAVLOGETPA 554  
 QY 512 GLRALDAAAOQOTSPEGVRSIADSSRGKLAFLFTGQAQTLGMGRGLYDVMSAFREAFD 571  
 DB 555 ELAELGLRAVAGSPARVARGLSTG-GRVGFLEFGQSQRLGMGRGLYDVMSAFREAFD 613  
 QY 572 -LCVRLFNQELDRPLREVMMAEPASVDALLDQRFQTPALFTFEYALAAAMRSGVPE 630  
 DB 614 EVCAQL-----DAVVDVSEELHRTGAAPALFAVEVALFRLLESGRAD 659  
 QY 631 LVAGSIGELVACVAGVSLDPAFLVAAGRLMQLPAGGAMVSIAPADVAALVAVP 690  
 DB 660 YVAGHSVGEALAAHVAGLSDDAKLVASARRLMQLPAGGAMVAVOATDE----VLP 715  
 QY 691 HAA-SVSTIAVNAVDQVITAGGCPVHAIAAAMAAAGARTALHVSASHPLMAPMIEA 749

DB 716 HLTDVAGIAINGRVSVGAEDVATALAEFRQGRKSSRLKVSHAFSLMDPLLEE 775  
 QY 750 FGRVAESVYRPSIVLVSNLGACTDEVSPPGYWRHAREVRFADGVYALAAAGCT 809  
 DB 776 FAEVYRLTFHRPQIPVVSNTGLA--EPTTPQYWRHAREVRFADGVYALAAAGCT 833  
 QY 810 FVEVPSKTLGLVPACMPD---ARPALASSRAGDEPATVLEALGGLMAVGLVSWAG 866  
 DB 834 FVEIPGCVLSALAEGLDDDVTVPAL---RADREPRVALVAVESLHAGVSPWHA 889  
 QY 867 LFPEGGRVPLPTPMQREVRITTDADAAGRRAPAG--HDEVEEGAGRGGR 923  
 DB 890 FFP-GARRVDLPYAFQERYWLD--ASPAPGCVRAAGLSADHPILGAASVLAGDER 946  
 QY 924 -SARLD--HP-----PES 934  
 DB 947 LUTRLSLRTHPMLSDNAVLCTALLPGTAFVELAVRAADENGALLDLELAEVLVPEQ 1006  
 QY 935 GRRE-----KVEAAGRPRL-----EIDEPVLHLVLTERRA----- 970  
 DB 1007 GGVAVQVWVGAVDGTGRPLTVHSRPEDDSDLPVW-RHATGVTEGSEBAGAAADRLLT 1065  
 QY 971 -PGLGEVELAVDAAGSFNDVOLALGM-----VPD--DLP 1002  
 DB 1066 AMPPTGSEPYDLD---GFYDRLAALGLAYGAPFRGLRSAMRTGEDVFAEVALPDGDTG 1121  
 QY 1003 GKRPPLLL-----GEGCAGRIVAVGEG----- 1025  
 DB 1122 SFLLHPALLDAALHAITAGGAGDSLVABADPPLFPAMSRVSATGASTRLVRLAPGT 1181  
 QY 1026 -VNGLV---VGQVVI-----ALSA-----GAFAT 1045  
 DB 1182 DAVSLVADAGEBVAVESITLRAVASERLKRSGDAMFTVERAPITLPSDDADGTAVA 1241  
 QY 1046 HTTTSAL-----VLPFQA-----LSAIEAAMPAYILT-AMYALDRILRL 1086  
 DB 1242 YVPLLAALAEADGSPQDVVWPCPDPEGATEAERAVASTDLRLVQRFADRPRL 1301  
 QY 1087 QPGERV--LHHAATGVGLAVQWAGVGAHVHATAGTPEKRAYL----- 1129  
 DB 1302 VLMARCDLHAHAAGLVRS--QAEHPGRVLTLETFRPDAALVLEGVLSGEPHYVR 1359  
 QY 1130 -----ESLGVRYVDSRSRDFVADVRAMTGEGVVYVLSLGEILDKSF--N 1175  
 DB 1360 EGEVASRLRAASLATEDEVATGRADENASALVGAASRLCTVLTGAAGL-GKTLARH 1418  
 QY 1176 LMRSHG-----RVELEKRCQYADNOGLAPFLRLNSF----- 1208  
 DB 1419 LVTHGVRRLVLSRRGADAPGADLKAELAAAGAEATWACDLAERDALRLAATPVD 1478  
 QY 1209 SLVDLRGMLE-----RPARVRAL-----LEELG-----LTAAGVFPFP 1244  
 DB 1479 SVHTAGVLDGCVATLTPEVGAVALPKADAVLNLDELGTPTTFLFSSAAGVFGNG 1538  
 QY 1245 IATLPIARV-ADAFRSNAQO----- 1264  
 DB 1539 QGVNAANAFDAARBARHQAQRTVSLANGLMGEGAMADABDAGSRMAAGVLP 1598  
 QY 1265 ----HLKVLTLGDE-VOIRLEPFA-----GAGSTGDRD 1296  
 DB 1599 TEDGLRFDALASNEEVLVPRLDLTAALRNRAETVPALRLGVVSLRTOARAAAGS 1658  
 QY 1297 LUDRLAAPARAALAEPLRTOVSOL--RTPEIKVGAELFTRLGMSLMAVELNRI 1355  
 DB 1659 LAERLAGLAADEREALLELRGVAAVLGHSSRAVOPTTAFODLGFDSLTAVELNRL 1718  
 QY 1356 EASIKLXLTSTFLSTSPNIALAONT-----DALATALSRLRVAAE----- 1397  
 DB 1719 ATATGLRLPATLVVDHTPALAHIGTELLEGEAPATAGAAAGTAADEPLAIVGMS 1778  
 QY 1398 -NLRAGVQND-----FVSSGAD-----QDEITIAL 1421  
 DB 1779 CRIPGCVRSPELWRLVASGRDGTTRPPDRMGVDGL 1816



Db 7590 DLVFAEVAAGIGHASDRTRVEHAFODIGFSDLTAVELRNLRNKATGRLPATLVFDPHT 7649  
Qy 1374 IALLAQLLDLALATA-----LSIERVAE---NLBAGVNDPFS 1410  
Db 7650 PALIAHVFMAAAGAEPAVDSLLADLDREBELVTTLAASEARDRLS 7699

## RESULT 10

US-10-156-761-7963  
; Sequence 7963, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7963  
; LENGTH: 1835  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7963

Query Match 29.8%; Score 2152; DB 15; Length 1835;  
Best Local Similarity 34.1%; Pred. No. 3e-156;  
Matches 609; Conservative 185; Mismatches 554; Indels 436; Gaps 49;

Qy 4 RPIRAEDPPIAIVGACRLPGVITDLSGFWTLLEGSDTIGRPAER-WDAAMPDPDP 62  
Db 25 RELFARAOEPIAIVGMSGRFPGVGSPPELRLLAEGEDALSQFPVRGMVDLYDPDP 84  
Qy 63 DAPKTPVTRASFSDVACFDASFFGISPREALRNDPARRLLLEVCWALEMAIAPSA 122  
Db 85 EOPKVTYTRGCGFLHDAADPAEFPGISPREALNDPQORLLLEAMEAFHAGIDPAG 144  
Qy 123 VGTETGVFTIGPSEY---EALPQATASAEIDAAGLGTMPSVGAGRISYALGRPCV 179  
Db 145 KGSRTGVFAGLMHYDGFPGVLPPEG---VEGLMSTGNSGSVASGRIVTFGEGPAV 199  
Qy 180 AVDTAVYSSSLVAHVLACOSLRSGECSTALAGVSLMSPSTLWLSKTRALARPGRCA 239  
Db 200 TVDTACSSSLVAHVLAVOALRTGCSMALAGVTVMGPFAVEFSRGRGLAPGRCSF 259  
Qy 240 SABADGGRGCGAVVVLKRLSGARADGRLIAVIRGSAINHDSGSLTPVNGSSOEIV 299  
Db 260 AAAADGASWSRGTLLEVERLRDARRLGKVLAVRGSAAVNOGASNGLTAPNGSQRV 319  
Qy 300 LKRALDAGCAASSVGVYEAHGTGTTGDPTEIGALNAVYLGVDVATPPLLIGSVKTN 359  
Db 320 ILLEALGARLSAEEDVAEAGTGTSLGDPTEAQAOLATYQGERE--RPLLLGSLKSNIG 377  
Qy 360 HPEVASGITGLLVLSLQHQIPAHVLAOLNPRISGDLRLVTYTRAKTWPMPNTPRR 419  
Db 378 HTQAAAGVAGVIAKVLMOHGVLPKTLHVPTRPRVMSVGSVLLTEALAMPETDHRP 437  
Qy 420 AGVSSFGMSTNAHVLEAPATCTPPAPER-----PAELL-----VTSARTAS 464  
Db 438 AGVSSFGISGNAHIIIEQAPPEAGDPADDTAGCEGVAEELIARKTIVVWVLSAKRGA 497  
Qy 465 ALDQAARLRDLHLETPSQCLGDVAFSLATTRSMERHLAVAAATSRGRLAALDAAGQ 524

Db 498 ALREQARLLSSVD-LAAASPADVGSLLATRSALREHRAAVGADGELRALITLAAAGE 556  
Qy 525 TSPCAVSIADSSRGKLAFLFTGGAQOTLGMGRGLVVMGAFREAF-DLCVRLFNQELDR 583  
Db 557 PAAGVTVGRAGADRGKGFLLFSGSGSORIGMELYAVVFPAAYNEVCALL--DL-- 611  
Qy 584 PLREVMMAEPASVDALLDQTAFTOPALFTFEYALALMRSWGEVLPVAGHSIGELVAA 643  
Db 612 -----FVDVDAETLHQGTSTOPTALFAMEVALFRLLESMAIRPDVYVAGHSIGEIAAA 662  
Qy 644 CVAGVFLEDAVFLVAARGLMQALPAGGMVSIIEAEADVAAAHAH-AASVSIIVANA 702  
Db 663 HVAAGVSLGDAAARIVSVRAALMQALPAGGMVAQAQAEDE---VLEYLGEVGAIAIING 718  
Qy 703 PDQVVIAGQOPVYIAIAAAGARFKALVSHAFSPMAFMLBAFGRVAESVSRAP 762  
Db 719 POSVYVSGAEDAVMAVAEVAAGRKTSRLKSHAFSPMLDPMLEFAAVRBLTGER 778  
Qy 763 SYLVSNLSGKACTDEVSPGVYWRHAREVVRPADGVKALHAAGAGTFVEVGPSTLLGL 822  
Db 779 RIVPVSNTLGRLA--EYTPHEHWVRHAREVVRPADGVKALHAAGAGTFVEVGPSTLLGL 836  
Qy 823 VPACMPDARPALAASSRAGDEPATVLEALGLMAVGLVSNAGLFPSSGRRVPLPTYPW 882  
Db 837 AOGCLDADIVTPALRTDRPEPOLTLVAGQLHTRGVSBDWQALFP-GALRYDLPTYAF 895  
Qy 883 QREERYWI--DTKADDAARG-----DRRAPG----- 905  
Db 896 QREERYWLKTTAVDVSSAGLSSGHALLGAVETAGQDVLFTGRSLPADLPTAERTVG 955  
Qy 906 -----AGHDEVE-----EGGAVR-----GGDRSA 925  
Db 956 GVPVLPVAAFEVLAASAGQAGFPYLEQLDVDAVLAEGCAVRIQVIGADDDPDRSV 1015  
Qy 926 RL-DHP-----PESGREGKYE-----AG 944  
Db 1016 RVSPRPAAGSGSEBWRHVTGKLRNAPAEFDMIMPFGAEBGDEVRGLRAWRAG 1075  
Qy 945 DRPF---RLRIDE-----PGVLDHLVLRTERA-----PG 972  
Db 1076 DELYAELELVTERDRADRALHPEVLDAALGMATSRBAADPGVLMFPMAGSGSVHTPG 1135  
Qy 973 LGEVEIIV-----DAAGLSFNDVQ-----LALGMVDD----- 1000  
Db 1136 PGNVRIRIRVGVDAVAIEISDADGSPVASVSRTRTMTAAQVLAARVAHQDMFEIDW 1195  
Qy 1001 -----LPGKRNPELL 1010  
Db 1196 VEHTAASAPATAVAADVAVLDEGVSHSIDALRKAVDGAATPAHVLLPCAPSEGDAAA 1255  
Qy 1011 LGGECAGRIYAVGCVNG-----LVNGOPVIALSAGAF-ATHVYTTSA---ALVLRP 1058  
Db 1256 TRGALAGVLSVVOQVADERPSSDRLVMTRGAEVLGTLDLAHVAVSGVLSAQTEHP 1315  
Qy 1059 QALSATBA-----AAMPVAYLTAWYAL-----RIARLQPGER- 1091  
Db 1316 DRLLLVADDSVTNGTGTVDGVDPALAALALADEQOAVRODKLVVPRHARTTPPEEP 1375  
Qy 1092 -----VLHAATGCVGLAAVOAHOYGAE-----VHAATAGTPKRAYL 1129  
Db 1376 TGHAMDEPGLVLLTGAAGGLGRSL--ARHIVVERGARRLILLASRRPAAEGVDELVAAL 1432  
Qy 1130 ESIG--VRVYSDRSRDE-FVADVRAWMTGCGVDVVLNSLSEGLLIXKSFNLLRSHGRVEL 1186  
Db 1433 SELGAHVDAACDIADDDALAEELASYPABEPLVAVVHAALDDGVITALTP----- 1485  
Qy 1187 GKRDY---ADNQLGLRPLRLNLSFS---LVDLRGM-----LERP 1221  
Db 1486 GRUDTVLRPRKADGALHHELTRDLNLSAFVLFSSLAIVLSAGAAVAAANAFDGLARQ 1545  
Qy 1222 ARVRAL-----LEELGLIAGVFTPPP 1244  
Db 1546 RRAAGLPSTLAWGLMTDNGMGMDRLTDADRNRMGAGVVPFSPABGLALFDALTVDRP 1605



```

; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Belknap, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3816
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
; OTHER INFORMATION: Oleandolide PKS
; US-09-808-880-3

Query Match      29.54; Score 2128; DB 11; Length 3816;
Best Local Similarity 28.04; Pred. No. 6.3e-154;
Matches 617; Conservative 216; Mismatches 544; Indels 828; Gaps 44;

QY 2 ADPPIERBA--EDPIAIVGASGRLPGGVIDLGSFWTLLEGSRDVTGVRPAER-WDAAM 57
DB 1581 ADLSVVRNADBEDDPVIVTGMACRFPGDIOTPEAFWTLLEGAGVISELPAANGMDMERL 1640
QY 58 FDDPDPAKGTPTVTRASFSLDVACFDSFPGISPREALRMDPAHLLLEVCEALENAAT 117
DB 1641 LNPDPKPKGTSAVRYGGFELYDAGEFDPAFPGISPREALRMDPOQRLLLETWELIESAGV 1700
QY 118 APSALVCTETGVFGIGPSEYEALPQATASAEIDAHGGLGTMPSVAGRSYALGRP 177
DB 1701 APSLSHRKRTGTGTIGSNGQFYAPLL--WNSGGLDEGTQGVGNAGSVNSGVAASLGEGR 1758
QY 178 CVAVDTPYSSSLVAVHLACSLRSSECTALAGVSLMLSPSTLWMLSKTRALARDGRCK 237
DB 1759 AVVVDTRACSSSLVLAHLAVALRGESLAIAAGVTWMTSTPDSVETRSRQGLSBDGRCK 1818
QY 238 AFSADADGFERGGCAVVLKRLSGARADDRILAVIRGSAINHAGSSGLTVPNGSSOE 297
DB 1819 AFASADADGFERGGCAVVLKRLSGARADDRILAVIRGSAINHAGSSGLTVPNGSSOE 1878
QY 298 IYVLRALADAGCAASSVGYEAHCTGTTLDPIEIOALNAVYGLGRVATPPLIGSKTN 357
DB 1879 RIVRRAALADAGLAPADVDVEAHGTGTRLDPIEAQALLATYQGRGGRPVVLGYSKN 1938
QY 358 LGHEPVAAGITGLKLVLSLHOGQIPALHQAOLNPRISW--GDLRLTVRARTPMPDMN 415
DB 1939 IGHQAAGAGVAGVWKVLALEGRVVPKTLVDESPHVDMSAGEVELAV--EAVPMRGG 1996
QY 416 TPRRAGVSSFGMSGTNAHVLEAPATCTPPAPERPAELL-----VLSARTASAL 466
DB 1997 RIVRAGVSSFGISGTNAHVLEAPATCTPPAPERPAELL-----EPEPEPTRVAAADLVVPMVSGRDAGAL 2052
QY 467 DAAARLRLHLETPYSCGLDVAFLATTSAMEHRLAVALATREGIALDLDAAGQTS 526
DB 2053 REDPAARLAHVSTGAGV--DVGMSLVATRSVPEHRAVWVGSSELDSAAESLAAGAGV 2111
QY 527 PGAVRSIADSSRKLAFLLTGGCAQTLGMRGLYDVVASAFREAFDLCTVRLFNQELDRPLR 586
DB 2112 PGVAVSGVAPREGRRVVFPPGQGSQWGMAGLLDACPVAEVAECAMVLDPTGMSLV 2171
QY 587 EVWMAEPASVDALLDQTAFTOPALFTFEYALALMWSKGVPELVAGHSIGELVAAQVA 646

```

```

DB 2172 EVL-----RGGEAVLGRVDVQDPALMAVMSLARTYRYGVPEPAVVGHSQGEIAACVA 2227
QY 647 GVPSLEDAVFLVARGRLMQALPAGGAMVSIPEADVAVAAPHAASVIAAVNAPDQV 706
DB 2228 GGLSLADGARVVLRLRRALARIAGGGGMSVGSASARVRMLTTYGRSVAAVANGBSST 2287
QY 707 VIAGAQPVHAIAAAMAAGARTKALVSHAFHSPLMAPMLEAFGRVAESVSYRPSIVL 766
DB 2288 VVGSDVQALDELLAGEREGRVARRVPVDYASHAQMDQLRDLLEALADITQHSVVF 2347
QY 767 VSNLSKACTDEVSISGTYVRRAREVRPADYKALHAAGAGTFVVGPSSTLLGLVPAC 826
DB 2348 FSTVTADWIDTTLADAGWFTMLRETVRQEAVEGLVAQMGAFVSCSPHVPVPGIEQT 2407
QY 827 MP--DARPLALASSRAGDEPA--TVLEALGLMAVGLSVNA-----GLPSSGRVPLP 878
DB 2408 LDALDQNAVLGSLR--RDEGGILRLTSLAEAFVQGVFVDMTHAEQMTF--RTYDLP 2462
QY 879 TYPQERRYW----- 888
DB 2463 TYPFORQHYWPKPAPAPAGANLGDVAVGLTAAQHPLLGAVEMPDSDGLVTGQISLRTH 2522
QY 889 ----- 888
DB 2523 PMLADHEVLGSLVLPQTAFVLELAVOADRAGYDVLDELTLLEALVLPDRGIQVRLALGP 2582
QY 889 -----IDTKADDA--RGDRAPG----- 905
DB 2583 SEADGRSLQLSHRPEBAAGFHRMTHASGFVVGCGTAARPTERPAGWPPAGAEPAALA 2642
QY 906 ----- 905
DB 2643 SDRYARLVENGYTYGPSFQGLHTWRHGGDVVAEVALPECTPADGYALHPALDAAVQAV 2702
QY 906 -----AGHD----- 910
DB 2703 GLGSFVEDPQVYLPFLMSDVTLLHATGATSLRVAVSPAGDVTVALALADAPAGVATGA 2762
QY 911 -----VEEGAV--RG--DRSARL----- 927
DB 2763 LRLRTTSAQOLARARSABEHAMFRVVEBGSAAADCRGAGGTTYEGEPAAGAAAGT 2822
QY 928 -----DHP--PES----- 934
DB 2823 WAVLGRVPAAVRTWQVDVVTALDTPDHPADPQSLADLAAGDTVDVVVTSLLSLASG 2882
QY 935 -----GREKEVA----- 943
DB 2883 ADEPLGNRRPPTAAEQDTAATVAGVASHALMALDLVQAWLADERRHTASRLVLTBRHAMTV 2942
QY 944 -----GDR----- 946
DB 2943 AESDPEPDLLEAFWGLVRSQAENPGRFVLADIDDEASWDALPRAVASAEVAIRAG 3002
QY 947 -----PFLRLEIDEPGLDHLVLRVTERRAPGLG--EVEIAYD 981
DB 3003 AVVVRPLARATBEGVLVADEAGPFWLDTVEAGTLANLALVPCPDASRLGPDEVIAYR 3062
QY 982 AAGLSFNDVOLALGMVPPDLLPGKRPPLLIGECAGRIYAVGEVGNLVVGOPIALISAG 1041
DB 3063 AAGVNFRRDVLALGMVPRDE-----GLMGAEAAVTVLEVGGVTTLLPAGDRVMGLVYG 3114
QY 1042 APATHTVTSAAVLPRPOLSAIEAAMPVAVYLTAAVLDRLARLPGRGRVLIHAATGV 1101
DB 3115 GGEPAVLTNHRMLVRRPKMSFAEASVPAFLTAIYALHDLGLRGESVLAHSAAGV 3174
QY 1102 GLAAVQMAQVGALEVAATAGTEPEKRAYLESAGV--RYVDSRSRDRFVADVAVATGEGVD 1159
DB 3175 GMAAVQALAHMALEVEVGT--SKGMDVLAAGQIGDEBHSSRTTEEGQFRAISGGRGID 3233
QY 1160 VVINSISGELIDKSFNLSRSHGRFVLEGRDCYADNQLG-----RPLRLNISFSFLVDLR 1214
DB 3234 VVINALSGDFVDSASALLREGGRFVEMGKTDIRTD--LGVVAGADVPIRIRYVAFDLAEG 3291

```

QY 1215 GMLERPARVALLBEILGLIAGVTPPTATIPARVADAFSMAQOHIGKLVLT- 1272  
Db 3292 A-ERIGQMLDEIMALFDAGVLRLLPLAMEVRRBAHEALRFVSOARIVGKVLTP 3345  
QY 1273 -----LGDPEVOIRI- 1282  
Db 3346 AALDAEGTVLTGAGTGLAVARHLVTEHDVRRLLVSRGVAPODLAEALGAEVVA 3405  
QY 1283 -----PTH-----AG----- 1287  
Db 3406 ACDVANRKAALKALIEDIPPEHPVTGIVHTAGVLDGVVSGLTPERVDTLKPXDAALLTL 3465  
QY 1288 -----AGPSTG----- 1293  
Db 3466 ESVGELDLDPALFYIFSSAASMLGPGQGSYAAANOPLDTLARHRRARGLTSVSLKGL 3525  
QY 1294 -----DRD-----LDRRL----- 1301  
Db 3526 WHEASGLTGLADIDRDMRSAGIAPMPTDEALHFDRAATELGDPVLLPMRLNEALEDR 3585  
QY 1302 -----ASAPAPAAALAEF-----LRTQV 1321  
Db 3586 AADGTLPELISGLVRRHRPSARAGTATAPATGPEAFAREIAAADPRRALRDVRRGHV 3645  
QY 1322 SQVL--RREIKVGAELFTSLGMSLMAVELRNRIEASLKLKLTFTFLSTSPNIALAQ 1379  
Db 3646 ALVIGHSGPE-AIDAEQAFRDIGPDSLTAVELRNNLNAETGRLRPGTLVPYPPDSALAD 3704  
QY 1380 NLMLALATA-----LSLRYAENLRAGVONDVYSSGADDD 1415  
Db 3705 HLELLAPATQPTAPLAELERV-EQLLSAASGCPASAVDEE 3748

RESULT 13  
US-09-861-289-2  
; Sequence 2, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D. H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438us1  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 5215  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-2

Query Match 29.5%; Score 2125; DB 10; Length 5215;  
Best Local Similarity 37.0%; Pred. No. 1.7e-153;  
Matches 581; Conservative 175; Mismatches 568; Indels 248; Gaps 46;

QY 1 VADRP1---ERAA--EDPIAIVGASCRLLPGVYIDLSGFMTLLGSRDVTGVRPAER-WDAA 55  
Db 2812 VAERASGCDRATDDPDAIVGMACKRYFGVSSPEDLRLVAEGTDAISEFPVNRGMWLE 2871  
QY 56 AMFPDDPDAPGKTPVTRASFLLSDVACPDASFPGISPREALRMDPAHRLLEVCWEALENA 115  
Db 2872 SLVDPPDSKGTTYCREGSGFLGAGDFDAFFGISPREALVMDPGORLLLEVMERALERA 2931  
QY 116 AIAPSALVTGTCTGFTIGPSEYEALPQATRAEIDAHGGL--GTMPVGAGRISVYLG 173  
Db 2932 GIDPSLSLRSRGVYVGAAGSGY--ASDPLVPEG---SEGYLITGSADAVMSGRISVYLG 2987

QY 174 LRGCVAVDTAAYSSVLAHVLACOSLSRGECSTALAGVGLMSPSTLYWLSKTRALARD 233  
Db 2988 LEEGSMTEVERPACSSSVLALHVALRHRGCGLAGVAVMDDPAFVFEFSRQKGLAAD 3047  
QY 234 GRCAFSAEADGFRGSGCAVVLKRLSGARADGRIILAVIRGSAINHDGASSGLTVPNG 293  
Db 3048 GRCAFSAADGTGMAGVGVLVLERLSDARRAGHTVIGLVGTGAIVQDASNGLTAPNG 3107  
QY 294 SSGEIVLKRLLADNAGCAASSGVVEAHGTGTTTGDPREIQALNAVYGLGDVATPILIGS 353  
Db 3108 PAQORVIAELADAGLSPEDVDAVEAGTGRIGDPIEAQALLAASGRNSGDHPLWLG 3167  
QY 354 VKTNLGPPEYASGITGLKVLSTLQHQI PAHLHAQALNRPISGCDRLTVTRARTWPD 413  
Db 3168 LKSNIGHAQAAGVGVYIKVLQALRHGLLPRTHADPPTHADSSGRVVLITSEVMQR 3227  
QY 414 WNTPRRAGVSSFGKSGTNAHVLEADPAATCTPPAE-----RPAE-----LVL 458  
Db 3228 TGRPRRTGVAFGGCTNAHVLEAPAE-----PPAPEPAGEAPGSRPAAGAGPLAMV 3283  
QY 459 SARTASALDQAARLRPHLETYPSCCGDVAFSLATTRSMEHRLVAATSRGLRALD 518  
Db 3284 SGRDEPLRQARLRDLHLSRTPGAPRPDIATSLAATRAAFDHRVAVLISDGAEALALD 3343  
QY 519 AAAGCOTSPGAIVSIADSSRGKLAFLFTGGCAOTLGMGRGLYDWSAFREAFDLCYRLFN 578  
Db 3344 ALAEGRGPAVAVGVRD--GRMAFLFTGGSGRAGAHDLHAHTFFAALDBVTDRLD 3402  
QY 579 QEIDRLPREVMAEPASVDALLDQTAFTOPALFTFEYALAAIWSGVEBELVAGHSIG 638  
Db 3403 PLIGRPIGALLDARPGSPAEALLDRTETYPALFAVVALHRLLEHGMKRPDILGHSVG 3462  
QY 639 ELVAACVAGVPSLEDAFVLYAARGLMQALPAGAMVSIAPREADVAAVAAPHAASIA 698  
Db 3463 ELAAHAHAGVLDLDDACALVAARGLMQRLPPGAMVSVRAGEDEVALLAGREDACVA 3522  
QY 699 AVNAPDOVVIAGAPPHAIATAAARAGARTKALHVSASHSPPLMABFAGRAESVS 758  
Db 3523 AVNGPRSVIVISGAEEVAEAAQAAGRGRTTRLRVAHAHSPLMQMLNGFRVVAAGLR 3582  
QY 759 YRPSIVLSNLGK-ACTDEVSSPGYVVRHAREVRFADGVKALHAAGTFVEVCPKS 817  
Db 3583 YRBEPLTVSVTVGRPARPELTGPDYVAQVEPVAFADAVTRAHLGARTFLETCPDG 3642  
QY 818 TLGLVYACMPDARPALASSRAGDEPAT-----VLEALGLMAYGLVSNAGLF 866  
Db 3643 VLCMAEECLEDDTVALLPAIH---KPGTAPHGPAPALGALRAAAAAYRGARVDMAGM 3698  
QY 869 PSG-----GRVPLPTYPWOREERYWIDTKADDAARGDRAG-----AGHDEVEE 913  
Db 3699 ADGPEGPARVELPVHAFRRHRYVL-----AFGRADVTDMMYRIGMDRLP- 3744  
QY 914 GGAVRGDDRARSAR--LDHPPE-----SGRREK-VEAAGRPFRFLEIDEGVLDHLVLRV 965  
Db 3745 --AVTGARFATAGRWLVTHPDSPRCRELSGHAERLRAAGASVPVLPVADAPAA--DRASFAA 3801  
QY 966 TERRAPGLG-EVEIADVAGISFNDVOLAIQMPVDDLPKGNPPLILGGEAGRIYAV-- 1022  
Db 3802 LIRSATPDTRGDTAAVAVAG-----LSL-LSEEDPFRHOAHDPVG-----VLATLS 3848  
QY 1023 -----SEGUNGVL--VGOPVIA-----LSAGAFATHTVTSNALVLP----- 1056  
Db 3849 LMQAMEEEAVEARWC/SRAVAADDERPVGAGALMGGRVAAALRPRRWGLVDLPA 3908  
QY 1057 RPOA--LSAIEAAMP-----VAYLTAM-YALDIRA-----LOGERVLIHAA 1097  
Db 3909 SPGAHMAAAVERIAGEDEDIATAVASGSGRILTRLPDDGGRTAABAYPRRGVLTATGG 3968  
QY 1098 TGVGLAIVQACQVAGAE-----VHAATGTPK--- 1125  
Db 3969 TGAIGHLARWLAAAGBHLLATSRKRPDAPGAAGLAEILLIGAKTTPAACDTADNDGL 4028  
QY 1126 -----RAYESLGVRYVS--DSRSDFVADVAM--TGEGGVADVVLNLSGE 1168



```
Db 4029 ARVLRALPEDTPTLAVFHAAGVGPVTPLSRTSPHFADVYAGKAAGAAHIDELTRELGAG 4088
Qy 1169 LIDKSFNLSHGRFVELGKRDYADNQLGLRPFLR----- 1204
Db 4089 L--DAFVLVSSGAGWGSAGGAGYAAANALDALARRAADGLPATSIANGWGGGCGMA 4146
Qy 1205 -NLSFSLVDLRGMLEERPARVRLLELLGLIAAGVTPPEPIAT-LPIARVADAFRSMAQ 1262
Db 4147 DEAGAEYLGRRGM---RPMAPVSALRAMATAIASG--EPCPTVHTDWERGEGFTAFRP 4201
Qy 1263 AQHLGKLVLTLDPEVOIRLPTHAAG-----PSTGRDLIDRLASAPARAALPAFL 1317
Db 4202 S-----PLIAGLGTPE-----GGGRAAETPEEGNATAADLTALPAELRTALELY 4247
Qy 1318 RTOVSQVLTPTPE-1KVGAELFTRLGMDSLMAVELARRIBASLTKLSTTFELSTSPNIAL 1376
Db 4248 RARTAAALGDDPRAVEAEGERFPMGFDLSATVLRKGLASATGLDLPDLLLPDRTPLA 4307
Qy 1377 LAONLIDALATA 1388
Db 4308 LAHLAEELATA 4319

RESULT 14
US-09-860-846-2
; Sequence 2, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikomycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860.846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105.537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-2

Query Match 29.5%; Score 2125; DB 10; Length 5215;
Best Local Similarity 37.0%; Pred. No. 1,7e-153;
Matches 581; Conservative 175; Mismatches 568; Indels 248; Gaps 46;

Qy 1 VADRP1---BRAA-EDPIATVGCSCRLPGVVIDLSGFTWLLSEGRDVGKVPAR-WDA 55
Db 2812 VAEGASFGGRADDDDDIATVAGMACRYGVSSPEDIMRLVAEGTDAISEPVRWGDLE 2871
Qy 56 AWPDPDADGKTPVTASFLSDVACFDASFGISPREALRMDPAHLLLEVCBALENA 115
Db 2872 SLVDPDESCKGTTCREGEFLGEGDFDAAFGISPREALVMDQGLLEVSREALERA 2931
Qy 116 AIPASALVGTETGVFIGISEVEYALPQATASAEIDAHGGL-GTWPVAGAGRISYALG 173
Db 2932 GIPPSIRSGRGVYGAHGSY-ASDPRLVPEG---SEGVLTLGSADAVMSGRISYALG 2987
Qy 174 LRBPQVAVDPYASSIYAVHLACOSTLSGEGSTLLAGVSLMSPLTWMLSKTRALARD 233
Db 2988 LEGSGMVEVTRACSSSLVALHLAVALLRHEGGLLAGVAVMADPAFVERSRQKGLAAD 3047
Qy 234 GRCKAFSAEADGFGRGBCAVVVKRLSGARADDRILIAVIRGSAIHHGSSGLTYPNG 293
Db 3048 GRCKAFSAADGTCMGAGVGVTLVERLSDRARAGHTVGLVGTGAIVODGASNGLTAPNG 3107
Qy 294 SSGEIVLKRALADGCAASSVGVYEAHGTGTTGDPLEIQALNNAVYGLGSDVATPLLIGS 353
```

```
Db 3108 PAQORVIAELADAGSPEDVDVAEAGTGTGLDPIEAGALLAASGRNSGDHPLWGS 3167
Qy 354 VKTNLGHPEYASGTTGLKVLVLSQGOIPAHLLAQLNLRISGMDRLTVTRARTMPD 413
Db 3168 LKSNIGHAQAAGVGVITKLOLRHGLRPTTHADEPTHADWSSGRVLLTSEVWQR 3227
Qy 414 WNTPRAGVSSFGMSGTNAHVLEAPPACTPPAPE-----RPAE-----LLVL 458
Db 3228 TGRPRRTGVSAFVGVTGNHVLEAPA-----PPAPEPAGEAGSRAAGAEGLMVMV 3283
Qy 459 SARTASLMDQAARLDHLETPSOCLGDVAFSLATTRSMERLVAATSRBGLAALD 518
Db 3284 SGRDEPALRQOARLDHLSRTGAPRDIATFSLAATRAAFDHRVAVLISDGAELAAALD 3343
Qy 519 AAAGGOTSPGAVSIDSRSKLAFLPTGGAGOTLIGRGGLYDOWSAFREAFDLCVLFN 578
Db 3344 ALMBGRGPVAVTGVRDRD-GRMAFLTGGSGQAGAHDLHAAHTFPASALDEVTBRLD 3402
Qy 579 QELDRPLEVYMAEPASVDAALLDQTAFTQPALFTEYALAAALWRSVGEPELVAGHSIG 638
Db 3403 PLIGRPLGALLDARPGSPBEALLDRTEYTOPALFAVEVALHRLLEHWGMRPDLLEGHSVG 3462
Qy 639 ELVAACVAGVFSIEDAVFLVAARGRLMQLPAGGAWSTIAPENDVAAVAAPHAASVIA 698
Db 3463 ELAAAHVAGVLDDDCALVAARGRLMQLRPPGAGAWSVRAGDEVRALLAGREDACVA 3522
Qy 699 AVNAPQVVIAGAGOPVHAATAAAMARGARTKALHVSASHSPPLMAFPLBAFGRVAVSVS 758
Db 3523 AVNGPSSVVISGAEVAEAAQAOLAGRGRTRLRLVAHAHSPMDGMLGFEVEVAGLR 3582
Qy 759 YRRPSIVLSNLSGK-ACTDEVSPGYWVHAAREVVRFADGVKALHAAGATPEVVGPKS 817
Db 3583 YREBELTVSVTGRPARPELGTQDYMVAQVEPRFADAVTAHRLGARTFLETGPDG 3642
Qy 818 TLIGLVPCPDPARPLLASSRAGDEPAT-----VLEALGIMAVGLVSNAGLP 868
Db 3643 VLCGMAEECLEDPTVALLPAIH---KPGTAPHPAPARPGALRAAAAYGARGARDWAGM 3698
Qy 869 PSG-----GRRVLPPTPMQREERYWIDTKDAAARGRRARG-----AGHDVEE 913
Db 3699 ADPREGPARRVELVHAFFHRRTWL-----ARGRAADTDMMYRIGMDRLP- 3744
Qy 914 GGAVRGGRDSAR--LDHPPE-----SGRREK-VEAAGDRPRLLEIDEGVLDHLVLRV 965
Db 3745 --AVTGGARTAGKWLVIHPSPRCRELSGHAEBALRAAGSVPPLPADARA-DRASFAA 3801
Qy 966 TERRAPGLG-EVEIADAGLSFNDOALGMVPDDIGKPNPPLIGECAGRIYAV-- 1022
Db 3802 LIRSATGTPDTRGDTAAVAGV-----LSL-LSEEDRPHRQHAVPVAG-----VLATLS 3848
Qy 1023 -----GEGVNGLV--VCGPVIA-----LSAGAFATHTTSAALVLP----- 1056
Db 3849 LQAMEBEAVEARWCSSRAVAADDERPVGAAALMGIRVAALEPRTRWGLVDLPA 3908
Qy 1057 RPOA---LSAIEAAMP-----VAYLTAM-YALDRITAR-----LOPGERVLTHAA 1097
Db 3909 SPQAAHMAAVERLAGEDQIIVRASGWSGRRLTLRLPDGCGRTAARPAVPRGVLTGCG 3968
Qy 1098 TGGVGLAAVQMAOVGA-----VHAATAGTPEK--- 1125
Db 3969 TGAIGGHLARWLAAAGEHIALTSRRGPDAPGAAGLEAELLIGAKVTFAACTADRDGL 4028
Qy 1126 -----RAYLESIGVRYVS--DSRSRPFADVAVAM--TGEGEVVVLNLSGE 1168
Db 4029 ARVLRALPEDTPTLAVFHAAGVGPVTPLSRTSPHFADVYAGKAAGAAHIDELTRELGAG 4088
Qy 1169 LIDKSFNLSHGRFVELGKRDYADNQLGLRPFLR----- 1204
Db 4089 L--DAFVLVSSGAGWGSAGGAGYAAANALDALARRAADGLPATSIANGWGGGCGMA 4146
Qy 1205 -NLSFSLVDLRGMLEERPARVRLLELLGLIAAGVTPPEPIAT-LPIARVADAFRSMAQ 1262
Db 4147 DEAGAEYLGRRGM---RPMAPVSALRAMATAIASG--EPCPTVHTDWERGEGFTAFRP 4201
```

1263 AOHIGKLVLTGDEVEOIRITTHAGAG-----PSTGDRDLDRLASAPARAALAEFL 1317  
Db 4202 S-----PLIAGIGTP-----GGRAAETPEEGNATAADLTALPPEALRTLRLELV 4247  
Qy 1318 RTQVSQVLRTPB-IKVGAELFTRLGMDSLMAVELRNRIEASLKUKLSTFPLSTSPNIAL 1376  
Db 4248 RARTAAALGLDDPAEVAEGERFPAMGFPDLSLATVRLRRGLASATGLDLPDLLFDRDTPAA 4307  
Qy 1377 LAONILDALATA 1388  
Db 4308 LAHLAEILATA 4319

RESULT 15  
US-09-988-384B-2  
; Sequence 2, Application US/09988384B  
; Publication No. US20030073824A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin.  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988,384B  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 2  
; LENGTH: 5215  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-988-384B-2

Query March 29.5%, Score 2125; DB 11; Length 5215;  
Best Local Similarity 37.0%, Pred. No. 1.7e-153;  
Matches 581; Conservative 175; Mismatches 568; Indels 248; Gaps 46;

Qy 1 VADRP---ERRA-EDPIAIVGASCRLLPGVYIDLSGFMTLEGSRDYVGRPAER-WDA 55  
Db 2812 VAEGASGCDATDDPIAIVGACRYGVSSPBDMLRLVAEGDALSFPVARKMDE 2871  
Qy 56 AWPDDDPAGKTPVTRASFSLDVACPDASFFGISPREALRMDPAHRLILEVCWALENA 115  
Db 2872 SLYPDDEPSKGTTCREGEFLEGAGDFDAFFGISPREALVMDPQRLILLEVSWEALERA 2931  
Qy 116 AIAESAVGTETGVFIGISPEVEALPQATASAEIDAHGGL--CTMPSVGAGRTSYALG 173  
Db 2932 GIDDSLSLGSRGVYVGAHGSY-ASDRLVPEG--SEGYLLTGSADAVMSGRISYALG 2987  
Qy 174 LRGCVAVDATYSSSLVAHLACOSLRSSECTALAGVSLMLSPSTLWMLSKTRALARD 233  
Db 2988 LEGSMYETKCSSSLVAHLAVALRHGEGCLALAGVAVMADPAALVEERSROKGLAAD 3047  
Qy 234 GRCAFSAEADDFGRGSCAVVTLKRLSGADGDRILAVIRGSAINHDASSGILTPNG 293  
Db 3048 GRCAFSMAADGTMAEGVGLVLERLSDARRAGHTVGLVGTAVNQDASNGILTPNG 3107  
Qy 294 SSOSITVLRALADGCAASSGVVEAHGTGTTIDPRIEIOALNVYGLGRDVAATPLIGS 353  
Db 3108 PAQGRVIAEALADGLSPEDDAVEAHGTGRLGDIPEAGALLAASGNRRSGDHPMLIGS 3167  
Qy 354 VKTNLGHEVYASGTLGLKVVLSLOHGOIPALHQAOLNPRISWMDLTVTYRARTPPD 413  
Db 3168 LKSNIGHAQAAGVIGVTKMLQALRHGILLPRTLHDEPTPAHDSNGVRLITSVPQR 3227  
Qy 414 WNTERRACVSSFGSGTNAHVLEAPAACTPPAE-----RPAE-----LLVL 458  
Db 3228 TGRBRRTGVSAFVGVTNAHVLEAPAE-----PPAPBAGAPGSGRAAEGEGLAVV 3283

Qy 459 SARTASALDAQAARLRDHLLETYPSCGLDVAFSLATTRSAMERHAAVATRSREGLRALD 518  
Db 3284 SGDEPRLRSQARRLRHLSTTPGARPRDIAFSLAATRAAFDRAVLIGSGALAAALD 3343  
Qy 519 AAGQTSFGAVRSIADSRGKLAFLFTGQAQTLGMRGLYDVWASFREAFDLCVRLFN 578  
Db 3344 ALAEGRGPAVVRVRD-GRMAFLFTGOOSQAGAAHDLHAHTEFFASALDEVTRLD 3402  
Qy 579 QELDPRLEVMWAEPPASVDALLDQTAFTQPALFTFEYALALMRSGVEPELVAGSISG 638  
Db 3403 PLIGRPLGALLDARPGSEPALDREYTOPALFAVEVALHRLLEHGMPEDDLGLGSHVG 3462  
Qy 639 ELVAACVAGVSEADAVFLVAARGLMQALPAGGAMVSIEADAVAAVAPHAASVIA 698  
Db 3463 ELAAHVAAGVLDLDDCALVAARGLMQRLPAGGAMVSVAAGEDEVALLAGREDACVA 3522  
Qy 699 AVNAPDOVITAGACQPVHAIAAAMAAAGARTKALHVAHSPPLMAEMLEAFGRVAVSVS 758  
Db 3523 AVNGPRSVITSGAEBAVAEAAQLAGRGRTRLRVAAHAFSPPLMDGMLAGFREVAAGLR 3582  
Qy 759 YRPSITLVSNLSGK-ACTDEVSPGYWVRHAREVVPFADGVKALHAAGCTFVEVPSKS 817  
Db 3583 YREBELTVSVTGTGRRPARPGBLTGPDYVAQVREPVAFDAVRAHRLGARTFLETGPDG 3642  
Qy 818 TLIGLVPAQMPDARPALASSRAGDEPAT-----VLEALGLMAVAGLVSMAGLF 868  
Db 3643 VLQMAEECELDVTALLPAIH---KPGTAHPGPAALPGALRAAAAAYGKARADMAQMH 3688  
Qy 869 PSG-----GRVPLFTYPMQERYWIDTKADDAARGDRARQ-----AGHDEVEE 913  
Db 3699 ADGPEGARRELPVNAFRHRRVYL-----AGRAADTDMMYRICMDRLP- 3744  
Qy 914 GGAIVRGDRSAR--LDHPRPE-----SGRREK-VEAAGDRPFLIEDDEGVLDHLVLRV 965  
Db 3745 --AVTGARKTAGRLVTHPDSPRCRLSGHAERLRAAGSPVLPVDAAPA--DRASFAA 3801  
Qy 966 TERRAPGLG-EVEIAVDAAGLSFNDVOLATGMVDDLPGRKNPPLILGGECAGRIVAV-- 1022  
Db 3802 LIRSATPDRTRGDTAAVAGV-----LSL-LSEDEPRHGOHAPVPAQ-----VLATLS 3848  
Qy 1023 -----GEGVNGLV--VGOPYIA-----LSAGAFTHVTTSAVLVP----- 1056  
Db 3849 LMQAMEEAAVEARWCVSRAVAADREPRVGAAGALMGRAVALRPRTRMGVLVDLPA 3908  
Qy 1057 RPQA---LSAIEAAMP-----VAYLTAW-YALDRIR-----LQGERVLIHAA 1097  
Db 3909 SPGAHMAAAVVERLAGEDDOIIVASGSGWRRLTRLPRDGGRTAADAYRPRGTVLVTGG 3968  
Qy 1098 TGGVGLAAVQMAQIVGAE-----VHAATGTPRK-- 1125  
Db 3969 TGAIGLHAKWLAAGAHEHLATLSRRGPDAPGAAGLAEELLGAKTTPAACOTDADNGL 4028  
Qy 1126 -----RAYLESIGVRVVS--DSRSDFVADVAM--TGSEGVVDVNLISGE 1168  
Db 4029 ARVLRAIRPDTPLTVAHFAAGVPQVTPLSRTSPREHFAVYVAGKAAGAHLDLRELGAG 4088  
Qy 1169 LIDKSPMLLSHGKRFVFLGKRDICYADNOLGIRPLR----- 1204  
Db 4089 L--DAFVLSSGAGVWSAGCAAGAAANAAALDALARRAADGLPATSIAWGWSGGMGCA 4146  
Qy 1205 -NLSFLVDIRGMWLEPARVALLEELGLIAAGVTPPIAT-LPIARVADAFRMAQ 1262  
Db 4147 DEAGAEYIGRRKM--RPMAPVSLRMAATRIASG--EPCPTVHTHMERFGEFTAFRP 4201  
Qy 1263 AOHIGKLVLTGDEVEOIRITTHAGAG-----PSTGDRDLDRLASAPARAALAEFL 1317  
Db 4202 S-----PLIAGIGTP-----GGRAAETPEEGNATAADLTALPPEALRTLRLELV 4247  
Qy 1318 RTQVSQVLRTPB-IKVGAELFTRLGMDSLMAVELRNRIEASLKUKLSTFPLSTSPNIAL 1376  
Db 4248 RARTAAALGLDDPAEVAEGERFPAMGFPDLSLATVRLRRGLASATGLDLPDLLFDRDTPAA 4307

Sat Oct 4 14:14:29 2003

us-10-014-717-2.rapb

Page 17

Qy	1377	LAQNLIDALATA	1388
		:	
Db	4308	LAHILAELIATA	4319

Search completed: October 2, 2003, 17:47:41  
Job time : 109.09 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:33 ; Search time 34.7527 Seconds  
(without alignments)  
1730.044 Million cell updates/sec

Title: US-10-014-717-2  
Perfect score: 7210  
Sequence: 1 VADRPPIERAAEDPIAIVGAS.....GVQNDPVSAGADQDWETIAL 1421

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pdp:\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pdp:\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pdp:\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pdp:\*
- 5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pdp:\*
- 6: /cgn2\_6/ptodata/1/iaa/Backfile1.pdp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7210	100.0	1421	US-09-335-409-2	Sequence 2, Appli
2	7210	100.0	1421	US-09-568-102-2	Sequence 2, Appli
3	7210	100.0	1421	US-09-567-969-2	Sequence 2, Appli
4	7210	100.0	1421	US-09-568-480-2	Sequence 2, Appli
5	7210	100.0	1421	US-09-568-486-2	Sequence 2, Appli
6	7210	100.0	1421	US-09-568-472-2	Sequence 2, Appli
7	7210	100.0	1421	US-09-567-899-2	Sequence 2, Appli
8	4137	57.4	2439	US-09-335-409-7	Sequence 7, Appli
9	4137	57.4	2439	US-09-568-102-7	Sequence 7, Appli
10	4137	57.4	2439	US-09-567-969-7	Sequence 7, Appli
11	4137	57.4	2439	US-09-568-480-7	Sequence 7, Appli
12	4137	57.4	2439	US-09-568-486-7	Sequence 7, Appli
13	4137	57.4	2439	US-09-568-472-7	Sequence 7, Appli
14	4137	57.4	2439	US-09-567-899-7	Sequence 7, Appli
15	3831.5	53.1	2259	US-09-413-814-70	Sequence 70, Appli
16	3311.5	45.9	7257	US-09-335-409-5	Sequence 5, Appli
17	3311.5	45.9	7257	US-09-568-102-5	Sequence 5, Appli
18	3311.5	45.9	7257	US-09-567-969-5	Sequence 5, Appli
19	3311.5	45.9	7257	US-09-568-480-5	Sequence 5, Appli
20	3311.5	45.9	7257	US-09-568-486-5	Sequence 5, Appli
21	3311.5	45.9	7257	US-09-568-472-5	Sequence 5, Appli
22	3311.5	45.9	7257	US-09-567-899-5	Sequence 5, Appli
23	2273	31.5	2152	US-09-036-987A-3	Sequence 3, Appli
24	2273	31.5	2152	US-09-370-700-3	Sequence 3, Appli
25	2273	31.5	2152	US-09-603-207-3	Sequence 3, Appli
26	2250	31.2	6095	US-09-144-085-2	Sequence 2, Appli
27	2245.5	31.1	2890	US-09-413-814-67	Sequence 67, Appli

28	2245.5	31.1	3798	3	US-09-335-409-6	Sequence 6, Appli
29	2245.5	31.1	3798	4	US-09-568-102-6	Sequence 6, Appli
30	2245.5	31.1	3798	4	US-09-567-969-6	Sequence 6, Appli
31	2245.5	31.1	3798	4	US-09-568-480-6	Sequence 6, Appli
32	2245.5	31.1	3798	4	US-09-568-486-6	Sequence 6, Appli
33	2245.5	31.1	3798	4	US-09-568-472-6	Sequence 6, Appli
34	2245.5	31.1	3798	4	US-09-567-899-6	Sequence 6, Appli
35	2194	30.4	5087	3	US-09-144-085-1	Sequence 1, Appli
36	2150	29.8	1841	2	US-08-804-227C-6	Sequence 6, Appli
37	2128	29.5	3816	3	US-09-428-517-3	Sequence 3, Appli
38	2127	29.5	3170	3	US-09-036-987A-4	Sequence 4, Appli
39	2127	29.5	3170	3	US-09-370-700-4	Sequence 4, Appli
40	2127	29.5	3170	4	US-09-603-207-4	Sequence 4, Appli
41	2125	29.5	5215	3	US-09-105-537-2	Sequence 2, Appli
42	2113	29.3	4928	3	US-09-036-987A-5	Sequence 5, Appli
43	2113	29.3	4928	3	US-09-370-700-5	Sequence 5, Appli
44	2113	29.3	4928	4	US-09-603-207-5	Sequence 5, Appli
45	2108.5	29.2	3562	4	US-09-679-279-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1									
US-09-335-409-2									
Sequence 2, Application US/09335409									
Patent No. 6121029									
GENERAL INFORMATION:									
APPLICANT: Schupp, Thomas									
APPLICANT: Ligon, James									
APPLICANT: Molnar, Istvan									
APPLICANT: Zitzke, Ross									
APPLICANT: Cyr, Devon									
APPLICANT: Goerlach, Joern									
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES									
FILE REFERENCE: 4-30582A									
CURRENT APPLICATION NUMBER: US/09/335,409									
CURRENT FILING DATE: 1999-06-17									
NUMBER OF SEQ ID NOS: 30									
SOFTWARE: Patentin Ver. 2.0									
SEQ ID NO 2									
LENGTH: 1421									
TYPE: PRT									
ORGANISM: Sorangium cellulosum									
US-09-335-409-2									
Query Match									
Best Local Similarity 100.0%; Score 7210; DB 3; Length 1421;									
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	VADRPPIERAAEDPIAIVGASCRLLPGVYIDLSGFWTLLGSRDPTVGRVPAERWDAAWFDP	60						
Db	1	VADRPPIERAAEDPIAIVGASCRLLPGVYIDLSGFWTLLGSRDPTVGRVPAERWDAAWFDP	60						
Qy	61	DDPAFGKTPYTRASFSLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENAAIAPS	120						
Db	61	DDPAFGKTPYTRASFSLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENAAIAPS	120						
Qy	121	ALVETGTVFTIGSPSEYEALPQATASAEIDAHGIGCTMPSCAGAIATALGRPCVA	180						
Db	121	ALVETGTVFTIGSPSEYEALPQATASAEIDAHGIGCTMPSCAGAIATALGRPCVA	180						
Qy	181	VPTAYSSSLVAVHLACQSLRSGRSTLACGVSIMLSPSTLVMLSKTRALARDRCFAFS	240						
Db	181	VPTAYSSSLVAVHLACQSLRSGRSTLACGVSIMLSPSTLVMLSKTRALARDRCFAFS	240						
Qy	241	AEADGFGRGCAVVLKRLSGARADDRILAVIRGSAINHDAASSGLTPVNGSSQEIYL	300						
Db	241	AEADGFGRGCAVVLKRLSGARADDRILAVIRGSAINHDAASSGLTPVNGSSQEIYL	300						
Qy	301	KRALADGCAASSGYVEAHGTGTTLDPIEIQALNNAVGLGRVATPLPLIGSKYTNLGH	360						
Db	301	KRALADGCAASSGYVEAHGTGTTLDPIEIQALNNAVGLGRVATPLPLIGSKYTNLGH	360						

361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRLTYTRARTPMPDMNTPPRA 420  
361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRLTYTRARTPMPDMNTPPRA 420  
421 GVSSFGMSGTNAHVLEBAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480  
421 GVSSFGMSGTNAHVLEBAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480  
481 PSOCIGDVAFSLATTRSMERHLAVALTSRGLRAALDAAAGQTSFGAVSISADSSRGK 540  
481 PSOCIGDVAFSLATTRSMERHLAVALTSRGLRAALDAAAGQTSFGAVSISADSSRGK 540  
541 LAFLFTGGAGOTLGMGRGLYDWSAFREAFDL CVRLFQEOELDRPLREVMMAEPASVDAL 600  
541 LAFLFTGGAGOTLGMGRGLYDWSAFREAFDL CVRLFQEOELDRPLREVMMAEPASVDAL 600  
601 LDQTAFTQPALFTFEYALALMRSMGVPELVAAGHSIGELVAAQVAGVSLDVAFLVAA 660  
601 LDQTAFTQPALFTFEYALALMRSMGVPELVAAGHSIGELVAAQVAGVSLDVAFLVAA 660  
661 RGRIMQALPAGGAMVSIAPADVAAVAAPHAASVSIAAVNAPOVVIAGAGQPVHAIAA 720  
661 RGRIMQALPAGGAMVSIAPADVAAVAAPHAASVSIAAVNAPOVVIAGAGQPVHAIAA 720  
721 AMAAGARTKALVSHAFSHPLMPLAFGRVAASVYSRRSITLVNLSKACTDEVS 780  
721 AMAAGARTKALVSHAFSHPLMPLAFGRVAASVYSRRSITLVNLSKACTDEVS 780  
781 SPGYVVRHAREVVRADGVKALHAAGAGTFVEVGEKSTLLGLVPACMPDARALLASSRA 840  
781 SPGYVVRHAREVVRADGVKALHAAGAGTFVEVGEKSTLLGLVPACMPDARALLASSRA 840  
841 GRDEPATVLEALGLIMAVGLVSMAGLPPSGGRVPLPTTYMQERBYWIDTYADDAAGD 900  
841 GRDEPATVLEALGLIMAVGLVSMAGLPPSGGRVPLPTTYMQERBYWIDTYADDAAGD 900  
901 RRAAPAGHDEVEEGAVRGDRRSARLDHPPESSRREVEAAGRPRLLEIDEGVLDH 960  
901 RRAAPAGHDEVEEGAVRGDRRSARLDHPPESSRREVEAAGRPRLLEIDEGVLDH 960  
961 LVLRATERRAPGLGEVEIAVDAAGLSFNDVOLAMVPDDLPCKENPPLLGGECAGRTV 1020  
961 LVLRATERRAPGLGEVEIAVDAAGLSFNDVOLAMVPDDLPCKENPPLLGGECAGRTV 1020  
1021 AVGEGVNGLVGQPVIALSAGAFATHVTTSALVLPPOALSAIEAAMPVAYLTAMATL 1080  
1021 AVGEGVNGLVGQPVIALSAGAFATHVTTSALVLPPOALSAIEAAMPVAYLTAMATL 1080  
1081 DRIARLOGERVLIHAATGGVGLAAVOMAHGAEVHATAGTPEKRAVLESIGVYVSDS 1140  
1081 DRIARLOGERVLIHAATGGVGLAAVOMAHGAEVHATAGTPEKRAVLESIGVYVSDS 1140  
1141 RSDREAVDVRAMTGGEGVDVNVNLSGELLIDKSFNILSRHGRFVLEGRKDCYADQGLR 1200  
1141 RSDREAVDVRAMTGGEGVDVNVNLSGELLIDKSFNILSRHGRFVLEGRKDCYADQGLR 1200  
1201 PFLRNLVSFLVDLRGMLEERPARVALLLELLGLIAAGVTPPIATLPIARVADAFRSM 1260  
1201 PFLRNLVSFLVDLRGMLEERPARVALLLELLGLIAAGVTPPIATLPIARVADAFRSM 1260  
1261 AOAQHLGLVLTLDGPEVOIRIPTHAGAGSTGDDDLDRLASAAPAAALAEFLRTO 1320  
1261 AOAQHLGLVLTLDGPEVOIRIPTHAGAGSTGDDDLDRLASAAPAAALAEFLRTO 1320  
1321 VSQVARTBEIKVGAALFTRLGMDSLMAVELNRIEASIKLSTTFSTPSNIALAQN 1380  
1321 VSQVARTBEIKVGAALFTRLGMDSLMAVELNRIEASIKLSTTFSTPSNIALAQN 1380  
1381 LLDALATALSLERVAENIRAGVQNDPVSSGADQDWEIATL 1421  
1381 LLDALATALSLERVAENIRAGVQNDPVSSGADQDWEIATL 1421

RESULT 2  
US-09-568-102-2  
; Sequence 2, Application US/09568102  
; Patent No. 6346404  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zitzke, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1421  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-102-2  
  
Query Match 100.0%; Score 7210; DB 4; Length 1421;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
1 VADRPBBAADPRAIVGASCRLLPGVYDLSGFMTLLGSRDTRGVRAERMDAAAFDP 60  
1 VADRPBBAADPRAIVGASCRLLPGVYDLSGFMTLLGSRDTRGVRAERMDAAAFDP 60  
61 DPDAPKTPVTRASFSLDPVACFDASFGISPREALRMDPARLLEVCMEALENAAIAPS 120  
61 DPDAPKTPVTRASFSLDPVACFDASFGISPREALRMDPARLLEVCMEALENAAIAPS 120  
121 ALVGTETGVFTIGIBSEYEALPQATSAEIDANGGLCTMPSVQAGRI SYALGLRCPVA 180  
121 ALVGTETGVFTIGIBSEYEALPQATSAEIDANGGLCTMPSVQAGRI SYALGLRCPVA 180  
181 VDTAYSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARDGRKAFS 240  
181 VDTAYSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARDGRKAFS 240  
241 AEADGFGRGECAYVVLKRLSGARADGDRILAVIRGSAINHGAASSGLTVPNGSSQETVL 300  
241 AEADGFGRGECAYVVLKRLSGARADGDRILAVIRGSAINHGAASSGLTVPNGSSQETVL 300  
301 KRALLADAGCAASSVGYEVAHGTGTTLDGPIETIOALNMYGUGRVATPPLLGSYKTNLGH 360  
301 KRALLADAGCAASSVGYEVAHGTGTTLDGPIETIOALNMYGUGRVATPPLLGSYKTNLGH 360  
361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRLTYTRARTPMPDMNTPPRA 420  
361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRLTYTRARTPMPDMNTPPRA 420  
421 GVSSFGMSGTNAHVLEBAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480  
421 GVSSFGMSGTNAHVLEBAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480  
481 PSOCIGDVAFSLATTRSMERHLAVALTSRGLRAALDAAAGQTSFGAVSISADSSRGK 540  
481 PSOCIGDVAFSLATTRSMERHLAVALTSRGLRAALDAAAGQTSFGAVSISADSSRGK 540  
541 LAFLFTGGAGOTLGMGRGLYDWSAFREAFDL CVRLFQEOELDRPLREVMMAEPASVDAL 600  
541 LAFLFTGGAGOTLGMGRGLYDWSAFREAFDL CVRLFQEOELDRPLREVMMAEPASVDAL 600  
601 LDQTAFTQPALFTFEYALALMRSMGVPELVAAGHSIGELVAAQVAGVSLDVAFLVAA 660  
601 LDQTAFTQPALFTFEYALALMRSMGVPELVAAGHSIGELVAAQVAGVSLDVAFLVAA 660

```

Qy 661 RGRIMQALPAGGAVNSIEAPEADVAAPAAASVSIANAAPQOVVIAGAGOPVHAIAA 720
Db 661 RGRIMQALPAGGAVNSIEAPEADVAAPAAASVSIANAAPQOVVIAGAGOPVHAIAA 720
Qy 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
Db 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
Qy 781 SPGYWVRHAREVRFADGVKALHAAAGACTFVEVGPCKSTLLGLVPACMPDAPALLASSRA 840
Db 781 SPGYWVRHAREVRFADGVKALHAAAGACTFVEVGPCKSTLLGLVPACMPDAPALLASSRA 840
Qy 841 GRDEPATVLEALGLMAVGLVSWAGLPFSGGRVPLPTYPWQERWIDTKADDAARGD 900
Db 841 GRDEPATVLEALGLMAVGLVSWAGLPFSGGRVPLPTYPWQERWIDTKADDAARGD 900
Qy 901 RRAAGAGHDEVEEGGAVRGDRRSARLDHPPESGREKVEAAGDRPRLEIDEPGLVDH 960
Db 901 RRAAGAGHDEVEEGGAVRGDRRSARLDHPPESGREKVEAAGDRPRLEIDEPGLVDH 960
Qy 961 LVLETERBARGLBEVEIAVDAAGLSFMDVOLALGMVDDLLPGKXNPPLLGGEGCAGRIV 1020
Db 961 LVLETERBARGLBEVEIAVDAAGLSFMDVOLALGMVDDLLPGKXNPPLLGGEGCAGRIV 1020
Qy 1021 AVGEVNGLVVGGOPVIALSAGAFTHVTSAAVLPRPOLSAIEAAMPVAVILTAWYAL 1080
Db 1021 AVGEVNGLVVGGOPVIALSAGAFTHVTSAAVLPRPOLSAIEAAMPVAVILTAWYAL 1080
Qy 1081 DRIARLQGEREVLIIHAATGVGLAAVQMAQHVGAENVHATAGTPEKRAYLESIGVRYVSDS 1140
Db 1081 DRIARLQGEREVLIIHAATGVGLAAVQMAQHVGAENVHATAGTPEKRAYLESIGVRYVSDS 1140
Qy 1141 RSDRFVADVRAWTGEGVDVVLNSISGELLTKSPTLLRSRGPVLELGRDCYADNOLGLR 1200
Db 1141 RSDRFVADVRAWTGEGVDVVLNSISGELLTKSPTLLRSRGPVLELGRDCYADNOLGLR 1200
Qy 1201 PFLNLSFSVLVDLGMMLERPARVARALBELLEGLIAAGVFTPPPIATLPIARVADAFRSM 1260
Db 1201 PFLNLSFSVLVDLGMMLERPARVARALBELLEGLIAAGVFTPPPIATLPIARVADAFRSM 1260
Qy 1261 AQAQHLGLVLTIGDPEVOIRIPTHAGAGPSTGDBDLDLASAPARAALAEAFRLTQ 1320
Db 1261 AQAQHLGLVLTIGDPEVOIRIPTHAGAGPSTGDBDLDLASAPARAALAEAFRLTQ 1320
Qy 1321 VSQVLRPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKLTSTFLSTSPNIALAQN 1380
Db 1321 VSQVLRPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKLTSTFLSTSPNIALAQN 1380
Qy 1381 LLDALATLALSLERVAENLRAGVQNDFFVSGADODMEITIAL 1421
Db 1381 LLDALATLALSLERVAENLRAGVQNDFFVSGADODMEITIAL 1421

```

```

RESULT 3
US-09-567-969-2
; Sequence 2, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligson, James
; APPLICANT: Molnar, Jettan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0

```

```

; SEQ ID NO 2
; LENGTH: 1421
; TYPE: PR
; ORGANISM: Sorangium cellulosum
US-09-567-969-2
Query Match 100.0%; Score 7210; DB 4; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 VADRPTEAAEDPFAIYAGSGLRPLGCVITDLSGFTLLRGSMDTYGRPAERMDAAAWDP 60
1 VADRPTEAAEDPFAIYAGSGLRPLGCVITDLSGFTLLRGSMDTYGRPAERMDAAAWDP 60
Qy 61 DPDAAGKTPVTRASFSLDVACFDSPFGISPREALRMDPAHRLLEVCMELENAATAPS 120
Db 61 DPDAAGKTPVTRASFSLDVACFDSPFGISPREALRMDPAHRLLEVCMELENAATAPS 120
Qy 121 ALVGTETGVFIGIPSEYEALPQATASAEIDAHGGLGTMPVAGARISVALGRPCVA 180
Db 121 ALVGTETGVFIGIPSEYEALPQATASAEIDAHGGLGTMPVAGARISVALGRPCVA 180
Qy 181 VDTAYSSSLVAVHILACQSLRSGECSTALAGVSLMLPSTLVMLSKTRALARDCRKAFS 240
Db 181 VDTAYSSSLVAVHILACQSLRSGECSTALAGVSLMLPSTLVMLSKTRALARDCRKAFS 240
Qy 241 AADGPRGECAYVVLKRLSGARADDDRILAVIRGSAINHDSGLTVPNSSQOEIVL 300
Db 241 AADGPRGECAYVVLKRLSGARADDDRILAVIRGSAINHDSGLTVPNSSQOEIVL 300
Qy 301 KRALADGCAASVGYEBAHGTGTLGDPTEIQALNAVYGGRDVATPLLIGSVKTNLGH 360
Db 301 KRALADGCAASVGYEBAHGTGTLGDPTEIQALNAVYGGRDVATPLLIGSVKTNLGH 360
Qy 361 PEVAGITGLIKVLSLQHQIPALHQAQINPRIISGDLRLVTRARTPMDNTPERRA 420
Db 361 PEVAGITGLIKVLSLQHQIPALHQAQINPRIISGDLRLVTRARTPMDNTPERRA 420
Qy 421 GVSSFGMSGTAHVLEAPATCTPPAPEPALVLSARTASALDAQAARLDHLETY 480
Db 421 GVSSFGMSGTAHVLEAPATCTPPAPEPALVLSARTASALDAQAARLDHLETY 480
Qy 481 PSQCLGVASFSLATTRSAMEHRLVAATSREGLRAALDAQAQOTSFGAVSIDSRSKG 540
Db 481 PSQCLGVASFSLATTRSAMEHRLVAATSREGLRAALDAQAQOTSFGAVSIDSRSKG 540
Qy 541 LAFLEFGQAGTGLMGRLYDWSAFREBAPDLCVRLFNQELDRPLREVMMAEPASVDAL 600
Db 541 LAFLEFGQAGTGLMGRLYDWSAFREBAPDLCVRLFNQELDRPLREVMMAEPASVDAL 600
Qy 601 LDQTAFTQPALFTEVEYLAALMRSGVPELVAGHSIGELVAAACVAFSLEDVFLVAA 660
Db 601 LDQTAFTQPALFTEVEYLAALMRSGVPELVAGHSIGELVAAACVAFSLEDVFLVAA 660
Qy 661 RGRIMQALPAGGAVNSIEAPEADVAAPAAASVSIANAAPQOVVIAGAGOPVHAIAA 720
Db 661 RGRIMQALPAGGAVNSIEAPEADVAAPAAASVSIANAAPQOVVIAGAGOPVHAIAA 720
Qy 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
Db 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
Qy 781 SPGYWVRHAREVRFADGVKALHAAAGACTFVEVGPCKSTLLGLVPACMPDAPALLASSRA 840
Db 781 SPGYWVRHAREVRFADGVKALHAAAGACTFVEVGPCKSTLLGLVPACMPDAPALLASSRA 840
Qy 841 GRDEPATVLEALGLMAVGLVSWAGLPFSGGRVPLPTYPWQERWIDTKADDAARGD 900
Db 841 GRDEPATVLEALGLMAVGLVSWAGLPFSGGRVPLPTYPWQERWIDTKADDAARGD 900
Qy 901 RRAAGAGHDEVEEGGAVRGDRRSARLDHPPESGREKVEAAGDRPRLEIDEPGLVDH 960
Db 901 RRAAGAGHDEVEEGGAVRGDRRSARLDHPPESGREKVEAAGDRPRLEIDEPGLVDH 960

```



QY 961 LVLRATERRAPGLGEVEIAVDAAGISFNDVOLALGMVDDLPCKNPPLLGGECAGRIY 1020  
DB 961 LVLRATERRAPGLGEVEIAVDAAGISFNDVOLALGMVDDLPCKNPPLLGGECAGRIY 1020  
QY 1021 AVGEVNGLVVGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080  
DB 1021 AVGEVNGLVVGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080  
QY 1081 DRIARLOPGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAYLESIGVYVSDS 1140  
DB 1081 DRIARLOPGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAYLESIGVYVSDS 1140  
QY 1141 RSDRFVAVRAMTGGEGVDVNLNSGELIDKSPULLSHGRFVELGRDCCADNOLGLR 1200  
DB 1141 RSDRFVAVRAMTGGEGVDVNLNSGELIDKSPULLSHGRFVELGRDCCADNOLGLR 1200  
QY 1201 PFLRLSFLVLDLGMMLERPARVALLLELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260  
DB 1201 PFLRLSFLVLDLGMMLERPARVALLLELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260  
QY 1261 AQAQHLGLVLTLDPEVOIRIPTHAGAPSTGDRDLDRLASAAPARAALBAFLRTQ 1320  
DB 1261 AQAQHLGLVLTLDPEVOIRIPTHAGAPSTGDRDLDRLASAAPARAALBAFLRTQ 1320  
QY 1321 VSQVARTBEIKVGAALPTRLGMDSLMAVELNRIEASIKLSTFTSTFNIALAQN 1380  
DB 1321 VSQVARTBEIKVGAALPTRLGMDSLMAVELNRIEASIKLSTFTSTFNIALAQN 1380  
QY 1381 LLDALATALLSERVAENLRAGVQNDPVSSGADQDMEITAL 1421  
DB 1381 LLDALATALLSERVAENLRAGVQNDPVSSGADQDMEITAL 1421

RESULT 4  
US-09-568-480-2  
Sequence 2, Application US/09568480

Patent No. 6355458  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zinkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,480  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1421  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-568-480-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRIEBAEDPIAIVGASCRLPGSVIIDLGSFTLLGSRDTCVRVPAERMDAAWDP 60  
DB 1 VADRIEBAEDPIAIVGASCRLPGSVIIDLGSFTLLGSRDTCVRVPAERMDAAWDP 60  
QY 61 DPDAKGTPTVTRASTLSDVACFDASFFGISPREALRMPDARHLLLEVCMEALENAAIAPS 120  
DB 61 DPDAKGTPTVTRASTLSDVACFDASFFGISPREALRMPDARHLLLEVCMEALENAAIAPS 120  
QY 121 ALVGTETGFIGIGSEYEALPOATASAEIDAHGGLGTMPSVGARISYALGRPCVA 180  
DB 121 ALVGTETGFIGIGSEYEALPOATASAEIDAHGGLGTMPSVGARISYALGRPCVA 180

DB 121 ALVGTETGFIGIGSEYEALPOATASAEIDAHGGLGTMPSVGARISYALGRPCVA 180  
QY 181 VDTAYSSSLVAVHLACOSLRSGEGSTALAGVSLMSTSTLVMISKTRTALARDGCKAFS 240  
DB 181 VDTAYSSSLVAVHLACOSLRSGEGSTALAGVSLMSTSTLVMISKTRTALARDGCKAFS 240  
QY 241 AEAQFGRGECQAVVYLKRLSGARADGRILAVIRGSAIINHDSAGSLTVPNGSQCEIVL 300  
DB 241 AEAQFGRGECQAVVYLKRLSGARADGRILAVIRGSAIINHDSAGSLTVPNGSQCEIVL 300  
QY 301 KRALADACAASSVGYVEAHGTGTTLDPIEIOALNAYVGLRDVATELLIGSVKTNLGH 360  
DB 301 KRALADACAASSVGYVEAHGTGTTLDPIEIOALNAYVGLRDVATELLIGSVKTNLGH 360  
QY 361 PEYASGITGLKVLSTLOHGOIPAHLAQALNPRISWGDLLTTRATTPMDWNTPPRA 420  
DB 361 PEYASGITGLKVLSTLOHGOIPAHLAQALNPRISWGDLLTTRATTPMDWNTPPRA 420  
QY 421 GVSSFGMSGTNAHVLEBAPAACTPPAPERPAELVTSARTASALDQAARLDRHLEBY 480  
DB 421 GVSSFGMSGTNAHVLEBAPAACTPPAPERPAELVTSARTASALDQAARLDRHLEBY 480  
QY 481 PSQCLGDVAFSLATTRSAMEHRLAVALTSREGLRAALDAAAQGTSPGAVSIADSSRGK 540  
DB 481 PSQCLGDVAFSLATTRSAMEHRLAVALTSREGLRAALDAAAQGTSPGAVSIADSSRGK 540  
QY 541 LAFLFTGGCAOTLGMKRGLYVWMSAFREAPDLCYRLFNOELDRPREVMAEPPASVDAL 600  
DB 541 LAFLFTGGCAOTLGMKRGLYVWMSAFREAPDLCYRLFNOELDRPREVMAEPPASVDAL 600  
QY 601 LDQAFTPALFTEFEYALALMRSGVPELVAGHSIGELVAAQVAFSLEDVFLVAA 660  
DB 601 LDQAFTPALFTEFEYALALMRSGVPELVAGHSIGELVAAQVAFSLEDVFLVAA 660  
QY 661 RGRLMQALPAGANVSIAPBADVAAVAAPPAASVIAAANVAPQVYIAGAGOPVHATA 720  
DB 661 RGRLMQALPAGANVSIAPBADVAAVAAPPAASVIAAANVAPQVYIAGAGOPVHATA 720  
QY 721 AMAARGARTKALVSHAFSPMLAMPLEAFBRVABSVYRPSIYLSNLSGKACTDEVS 780  
DB 721 AMAARGARTKALVSHAFSPMLAMPLEAFBRVABSVYRPSIYLSNLSGKACTDEVS 780  
QY 781 SPGVVRRHAREVVRPADGVKALHAAAGTFVEVBPKSTLLGLVPAQMPDARPALIASRA 840  
DB 781 SPGVVRRHAREVVRPADGVKALHAAAGTFVEVBPKSTLLGLVPAQMPDARPALIASRA 840  
QY 841 GRDEPATVLEALGGLMAVGLVSWAGLFPSSGRRVPLPTYPMQBERYVIDTKADDAAGD 900  
DB 841 GRDEPATVLEALGGLMAVGLVSWAGLFPSSGRRVPLPTYPMQBERYVIDTKADDAAGD 900  
QY 901 RRAAGAGHDEVEEGANVGDRRSARLDHPPESGRKRVNAAGDRPRLTIDEFGVLDH 960  
DB 901 RRAAGAGHDEVEEGANVGDRRSARLDHPPESGRKRVNAAGDRPRLTIDEFGVLDH 960  
QY 961 LVLRTERRAPGLGEVEIAVDAAGISFNDVOLALGMVDDLPCKNPPLLGGECAGRIY 1020  
DB 961 LVLRTERRAPGLGEVEIAVDAAGISFNDVOLALGMVDDLPCKNPPLLGGECAGRIY 1020  
QY 1021 AVGEVNGLVVGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080  
DB 1021 AVGEVNGLVVGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080  
QY 1081 DRIARLOPGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAYLESIGVYVSDS 1140  
DB 1081 DRIARLOPGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAYLESIGVYVSDS 1140  
QY 1141 RSDRFVAVRAMTGGEGVDVNLNSGELIDKSPULLSHGRFVELGRDCCADNOLGLR 1200  
DB 1141 RSDRFVAVRAMTGGEGVDVNLNSGELIDKSPULLSHGRFVELGRDCCADNOLGLR 1200  
QY 1201 PFLRLSFLVLDLGMMLERPARVALLLELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260  
DB 1201 PFLRLSFLVLDLGMMLERPARVALLLELLGLIAAGVFTPPPIATLPIARVADAFRSM 1260

QY 1261 AQAQHLGKLVLTGDEPVOQIRIPTHAGAPSTGDRDLDRLASAAPARAALAEFLRTQ 1320  
DB 1261 AQAQHLGKLVLTGDEPVOQIRIPTHAGAPSTGDRDLDRLASAAPARAALAEFLRTQ 1320  
QY 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380  
DB 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380  
QY 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421  
DB 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421

RESULT 5  
US-09-568-486-2  
; Sequence 2, Application US/09568486  
; Patent No. 6358459  
; GENERAL INFORMATION:  
; APPLICANT: Schnupp, Thomas  
; APPLICANT: Ligoun, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,486  
; PRIOR FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1421  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-486-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRPRIERAEDPIAIVGASCRIPGVYIDLSGFMTLLGSRDVTGRVPAERMDAAAFDP 60  
DB 1 VADRPRIERAEDPIAIVGASCRIPGVYIDLSGFMTLLGSRDVTGRVPAERMDAAAFDP 60  
QY 61 DDDAPGKTPVTRASFSLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENAAIAPS 120  
DB 61 DDDAPGKTPVTRASFSLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENAAIAPS 120  
QY 121 ALVGTGTVFTIGSPSEYEALPQATSAETIDAHGIGTWPVAGRISTALGIRGCVA 180  
DB 121 ALVGTGTVFTIGSPSEYEALPQATSAETIDAHGIGTWPVAGRISTALGIRGCVA 180  
QY 181 VDTAVSSSLVAVHLACOSLRSGECSTALAGVSLMSPSTLWMLSKTRALARDRCAPFS 240  
DB 181 VDTAVSSSLVAVHLACOSLRSGECSTALAGVSLMSPSTLWMLSKTRALARDRCAPFS 240  
QY 241 AADGFRGEGCAVVLKRLSGARADDRILAVIRGSAINHOGASSGLTVPNSSQEIIVL 300  
DB 241 AADGFRGEGCAVVLKRLSGARADDRILAVIRGSAINHOGASSGLTVPNSSQEIIVL 300  
QY 301 KRALADAGCAASVGYEAHGTGTTLDPIEIOALNVYGVGRVATPPLIGVKTMLGH 360  
DB 301 KRALADAGCAASVGYEAHGTGTTLDPIEIOALNVYGVGRVATPPLIGVKTMLGH 360  
QY 361 PEVASGITGLKLVLSLQHGQIPAHLAQALNPRISWGDRLITVTRARTPMDNTPRRA 420  
DB 361 PEVASGITGLKLVLSLQHGQIPAHLAQALNPRISWGDRLITVTRARTPMDNTPRRA 420  
QY 421 GVSSFGSGTNAHVLEAPPAATCTPPAPERPAELVLVSARTASALDQAARLDHLETY 480  
DB 421 GVSSFGSGTNAHVLEAPPAATCTPPAPERPAELVLVSARTASALDQAARLDHLETY 480

DB 421 GVSSFGSGTNAHVLEAPPAATCTPPAPERPAELVLVSARTASALDQAARLDHLETY 480  
QY 481 PSOCIGVAFSLATTSAMEHRLAVALATSREGRAALDAAQOCTSPGAVRSIADSSRGK 540  
DB 481 PSOCIGVAFSLATTSAMEHRLAVALATSREGRAALDAAQOCTSPGAVRSIADSSRGK 540  
QY 541 LAELFTGQAGTQAGMGLVDVMSAFREAFDLCVRLPNOELDRPLREVMMAEPASVDAL 600  
DB 541 LAELFTGQAGTQAGMGLVDVMSAFREAFDLCVRLPNOELDRPLREVMMAEPASVDAL 600  
QY 601 LDQTAFTQPALFTFEYALALMSWGPBELVAGHSIGELVAAVCVAVFSLQDVFVAA 660  
DB 601 LDQTAFTQPALFTFEYALALMSWGPBELVAGHSIGELVAAVCVAVFSLQDVFVAA 660  
QY 661 RGRIMQALPAGAMVSTIAPADVAALVAPHAASVSTIAVNAPOQVVIAGAGOVHAIIA 720  
DB 661 RGRIMQALPAGAMVSTIAPADVAALVAPHAASVSTIAVNAPOQVVIAGAGOVHAIIA 720  
QY 721 AMARQARTKALHVSHPFMAFMLAEFGRVAVSVYRPSITVLVSNLSGKACTDEVS 780  
DB 721 AMARQARTKALHVSHPFMAFMLAEFGRVAVSVYRPSITVLVSNLSGKACTDEVS 780  
QY 781 SPGYWRHAEVVRPADGVYALHAAGTVEVGPSTLLGLVPACMPDARPLLASRA 840  
DB 781 SPGYWRHAEVVRPADGVYALHAAGTVEVGPSTLLGLVPACMPDARPLLASRA 840  
QY 841 GRDEPATVLEALGLMAVGLVSMAGLFPSSGRRVPLPTYPWOBERRYIDTKADDAARGD 900  
DB 841 GRDEPATVLEALGLMAVGLVSMAGLFPSSGRRVPLPTYPWOBERRYIDTKADDAARGD 900  
QY 901 RRAFGAGHDEVEEGGAVRGDRRSARLDHPPPSGRRREKVEAAGDRPFLIEDEPGVLDH 960  
DB 901 RRAFGAGHDEVEEGGAVRGDRRSARLDHPPPSGRRREKVEAAGDRPFLIEDEPGVLDH 960  
QY 961 LVLRTERRAPGIGVEEIAVDAAGLSFNDVQLAGVPPDLPGKPNPPLLGGCAGRIIV 1020  
DB 961 LVLRTERRAPGIGVEEIAVDAAGLSFNDVQLAGVPPDLPGKPNPPLLGGCAGRIIV 1020  
QY 1021 AVEGVNGLVVGOPIALASGAFATHTTSAALVLPPOALSAIEAAMFVAYLTAVYAL 1080  
DB 1021 AVEGVNGLVVGOPIALASGAFATHTTSAALVLPPOALSAIEAAMFVAYLTAVYAL 1080  
QY 1081 DRIARLOPGERVLIHAATGVGGLAAVQMAQHVGAENVATGTEPKRAYLESIGVRYVSDS 1140  
DB 1081 DRIARLOPGERVLIHAATGVGGLAAVQMAQHVGAENVATGTEPKRAYLESIGVRYVSDS 1140  
QY 1141 RSDRFVADVAVMTGGEGVDVVLNSLSEGLIDKSFNLRSHGRFVELEGKRCYADNOLGLR 1200  
DB 1141 RSDRFVADVAVMTGGEGVDVVLNSLSEGLIDKSFNLRSHGRFVELEGKRCYADNOLGLR 1200  
QY 1201 PFLRNLSFSLVDLRGMWLEBPVRLALBELGLIAAGVTPPIATLPIARVADARSM 1260  
DB 1201 PFLRNLSFSLVDLRGMWLEBPVRLALBELGLIAAGVTPPIATLPIARVADARSM 1260  
QY 1261 AQAQHLGKLVLTGDEPVOQIRIPTHAGAPSTGDRDLDRLASAAPARAALAEFLRTQ 1320  
DB 1261 AQAQHLGKLVLTGDEPVOQIRIPTHAGAPSTGDRDLDRLASAAPARAALAEFLRTQ 1320  
QY 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380  
DB 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380  
QY 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421  
DB 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421

RESULT 6  
US-09-568-472-2  
; Sequence 2, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schnupp, Thomas

APPLICANT: Ligon, James  
 APPLICANT: Molnar, Istvan  
 APPLICANT: Zirkle, Ross  
 APPLICANT: Cyr, Devon  
 APPLICANT: Goerlach, Joern  
 TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 FILE REFERENCE: 4-30582A  
 CURRENT APPLICATION NUMBER: US/09/567,472  
 CURRENT FILING DATE: 2000-05-10  
 PRIOR APPLICATION NUMBER: 09/335,409  
 PRIOR FILING DATE: 1999-06-17  
 NUMBER OF SEQ ID NOS: 30  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO 2  
 LENGTH: 1421  
 TYPE: PRT  
 ORGANISM: Sorangium cellulosum  
 US-09-567-472-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRIERAADPIAIVGASCRIPGVITLGGFTLLGSRDVTGVRPAERMDAAAFDP 60  
 DB 1 VADRIERAADPIAIVGASCRIPGVITLGGFTLLGSRDVTGVRPAERMDAAAFDP 60  
 QY 61 DPDAKPTVTRASFSLDVACFDASFEGISPREALRMDPAHRLLEVCMEALENNAIAPS 120  
 DB 61 DPDAKPTVTRASFSLDVACFDASFEGISPREALRMDPAHRLLEVCMEALENNAIAPS 120  
 QY 121 ALVGETGVFIIGIGSEYEALPQATSAEIDAHGGLCTMPSVGGRISYALGRGPCVA 180  
 DB 121 ALVGETGVFIIGIGSEYEALPQATSAEIDAHGGLCTMPSVGGRISYALGRGPCVA 180  
 QY 121 ALVGETGVFIIGIGSEYEALPQATSAEIDAHGGLCTMPSVGGRISYALGRGPCVA 180  
 DB 121 ALVGETGVFIIGIGSEYEALPQATSAEIDAHGGLCTMPSVGGRISYALGRGPCVA 180  
 QY 181 VDTAVSSSLVAVHLACQSLRSCECTALAGVSLMSPSTLWLSKTRTALARDGRCKAFS 240  
 DB 181 VDTAVSSSLVAVHLACQSLRSCECTALAGVSLMSPSTLWLSKTRTALARDGRCKAFS 240  
 QY 241 AADGFGEGCAVVTLRKLSGARADGRITLAVIRGSAIINHDASSGLTPNGSSQEIYVL 300  
 DB 241 AADGFGEGCAVVTLRKLSGARADGRITLAVIRGSAIINHDASSGLTPNGSSQEIYVL 300  
 QY 241 AADGFGEGCAVVTLRKLSGARADGRITLAVIRGSAIINHDASSGLTPNGSSQEIYVL 300  
 DB 241 AADGFGEGCAVVTLRKLSGARADGRITLAVIRGSAIINHDASSGLTPNGSSQEIYVL 300  
 QY 301 KRALADACCAASVGVYEAHGTGTPPIEIOANAYVGRDVAETPLIGSVYTNLGH 360  
 DB 301 KRALADACCAASVGVYEAHGTGTPPIEIOANAYVGRDVAETPLIGSVYTNLGH 360  
 QY 301 KRALADACCAASVGVYEAHGTGTPPIEIOANAYVGRDVAETPLIGSVYTNLGH 360  
 DB 301 KRALADACCAASVGVYEAHGTGTPPIEIOANAYVGRDVAETPLIGSVYTNLGH 360  
 QY 361 PEYASGITGLKVVLSLQHQI PAHIAAOALNPRISMGLRLTVTRARTPMPDMWTPRRA 420  
 DB 361 PEYASGITGLKVVLSLQHQI PAHIAAOALNPRISMGLRLTVTRARTPMPDMWTPRRA 420  
 QY 421 GVSFSGSGTAAHVLEBAPATCTPPAPERPAELIVSARTASALDQAARLRDHLTTY 480  
 DB 421 GVSFSGSGTAAHVLEBAPATCTPPAPERPAELIVSARTASALDQAARLRDHLTTY 480  
 QY 421 GVSFSGSGTAAHVLEBAPATCTPPAPERPAELIVSARTASALDQAARLRDHLTTY 480  
 DB 421 GVSFSGSGTAAHVLEBAPATCTPPAPERPAELIVSARTASALDQAARLRDHLTTY 480  
 QY 481 PSOCIGDVAFSLATTRSMERHLAAATSRGLRALDAAQCGTSPCAVSIADSSSGK 540  
 DB 481 PSOCIGDVAFSLATTRSMERHLAAATSRGLRALDAAQCGTSPCAVSIADSSSGK 540  
 QY 541 LAFLFTGGAOTLGMGRGLYDVWSAFREAFDLCVLFNOELDRPLREVMMAEPASVDAL 600  
 DB 541 LAFLFTGGAOTLGMGRGLYDVWSAFREAFDLCVLFNOELDRPLREVMMAEPASVDAL 600  
 QY 541 LAFLFTGGAOTLGMGRGLYDVWSAFREAFDLCVLFNOELDRPLREVMMAEPASVDAL 600  
 DB 541 LAFLFTGGAOTLGMGRGLYDVWSAFREAFDLCVLFNOELDRPLREVMMAEPASVDAL 600  
 QY 601 LDQTAFTOPALFTFEYALAAALMRSGVBEPELVAGHSISELVAAACVAGFSLDEDAFLVAA 660  
 DB 601 LDQTAFTOPALFTFEYALAAALMRSGVBEPELVAGHSISELVAAACVAGFSLDEDAFLVAA 660  
 QY 601 LDQTAFTOPALFTFEYALAAALMRSGVBEPELVAGHSISELVAAACVAGFSLDEDAFLVAA 660  
 DB 601 LDQTAFTOPALFTFEYALAAALMRSGVBEPELVAGHSISELVAAACVAGFSLDEDAFLVAA 660  
 QY 661 RGRIMQALPAGAMVSIAPEDVAAVAPHAASVIAAVNAPDVOVLAGQPHALAA 720  
 DB 661 RGRIMQALPAGAMVSIAPEDVAAVAPHAASVIAAVNAPDVOVLAGQPHALAA 720  
 QY 721 AMAARGARTKALHVSFAHSPILMAPLEAFGRVAESVYRRPSIYLVNLSGKACTDEVS 780  
 DB 721 AMAARGARTKALHVSFAHSPILMAPLEAFGRVAESVYRRPSIYLVNLSGKACTDEVS 780

DB 721 AMAARGARTKALHVSFAHSPILMAPLEAFGRVAESVYRRPSIYLVNLSGKACTDEVS 780  
 QY 781 SPGIWVRHAREVVFAPGVKALHAAGAGTFVEVCPKSTLLGLVAPCHPDPARPALLASSRA 840  
 DB 781 SPGIWVRHAREVVFAPGVKALHAAGAGTFVEVCPKSTLLGLVAPCHPDPARPALLASSRA 840  
 QY 841 GRDEPATVLEALGLMAVGLSVNAGLFPSCGRVPPPTYPMQERWIDTKADDAARGD 900  
 DB 841 GRDEPATVLEALGLMAVGLSVNAGLFPSCGRVPPPTYPMQERWIDTKADDAARGD 900  
 QY 841 GRDEPATVLEALGLMAVGLSVNAGLFPSCGRVPPPTYPMQERWIDTKADDAARGD 900  
 DB 841 GRDEPATVLEALGLMAVGLSVNAGLFPSCGRVPPPTYPMQERWIDTKADDAARGD 900  
 QY 901 RRABAGHDEVEEGCAVVGDRSARLDHPPESSGREKVEAAGDRPRLIDEPGVLDH 960  
 DB 901 RRABAGHDEVEEGCAVVGDRSARLDHPPESSGREKVEAAGDRPRLIDEPGVLDH 960  
 QY 901 RRABAGHDEVEEGCAVVGDRSARLDHPPESSGREKVEAAGDRPRLIDEPGVLDH 960  
 DB 901 RRABAGHDEVEEGCAVVGDRSARLDHPPESSGREKVEAAGDRPRLIDEPGVLDH 960  
 QY 961 LVLRTERRAAGLGEVEIYDAAAGLSTNDVQALAGMPPDDLPGKPNPPLLGGECARIV 1020  
 DB 961 LVLRTERRAAGLGEVEIYDAAAGLSTNDVQALAGMPPDDLPGKPNPPLLGGECARIV 1020  
 QY 961 LVLRTERRAAGLGEVEIYDAAAGLSTNDVQALAGMPPDDLPGKPNPPLLGGECARIV 1020  
 DB 961 LVLRTERRAAGLGEVEIYDAAAGLSTNDVQALAGMPPDDLPGKPNPPLLGGECARIV 1020  
 QY 1021 AVGEGVNGLVVGGQVIALSACAFTHVTTSAALVLPQQALSAIEAAMPVAYLTAWYAL 1080  
 DB 1021 AVGEGVNGLVVGGQVIALSACAFTHVTTSAALVLPQQALSAIEAAMPVAYLTAWYAL 1080  
 QY 1081 DRIARLOPGERVLIHAAATGVGLAAVQAOHVGAENATAGTPPKRAVLESIGVRYVSDS 1140  
 DB 1081 DRIARLOPGERVLIHAAATGVGLAAVQAOHVGAENATAGTPPKRAVLESIGVRYVSDS 1140  
 QY 1141 RSDRFVADVRATWGTGEGVDVYVNLSELIDKSNLRSRFRFELGKDCYADNOGLR 1200  
 DB 1141 RSDRFVADVRATWGTGEGVDVYVNLSELIDKSNLRSRFRFELGKDCYADNOGLR 1200  
 QY 1201 PFLNLSFSLVDLGMMLERPARVALLLEGLIAGVFTPPPIATLPIARVADAFRSM 1260  
 DB 1201 PFLNLSFSLVDLGMMLERPARVALLLEGLIAGVFTPPPIATLPIARVADAFRSM 1260  
 QY 1261 AQAHLGKLVTLTCDPEVOIRIPTHAGAGPSTGDRDLDRILASAPARAALAEFLRTQ 1320  
 DB 1261 AQAHLGKLVTLTCDPEVOIRIPTHAGAGPSTGDRDLDRILASAPARAALAEFLRTQ 1320  
 QY 1321 VSQVLRPEIRVGAELFTRIGMDSIMAVEIRNIEASLKKISTFPLSTSPNIALAQN 1380  
 DB 1321 VSQVLRPEIRVGAELFTRIGMDSIMAVEIRNIEASLKKISTFPLSTSPNIALAQN 1380  
 QY 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDWEIHAL 1421  
 DB 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDWEIHAL 1421

RESULT 7  
 US-09-567-899-2  
 ; Sequence 2, Application US/09567899  
 ; Patent No. 6383787  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Zirkle, Ross  
 ; APPLICANT: Cyr, Devon  
 ; APPLICANT: Goerlach, Joern  
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 ; FILE REFERENCE: 4-30582A  
 ; CURRENT APPLICATION NUMBER: US/09/567, 899  
 ; CURRENT FILING DATE: 2000-05-10  
 ; PRIOR APPLICATION NUMBER: 09/335,409  
 ; PRIOR FILING DATE: 1999-06-17  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: Patent In Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 1421  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 US-09-567-899-2  
 Query Match 100.0%; Score 7210; DB 4; Length 1421;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 VADRIEAAEDPIAIVGASCRLLPGGVIDLGGFWTLLEGSRTVGRVPAERMDAAAMPDP 60
Db 1 VADRIEAAEDPIAIVGASCRLLPGGVIDLGGFWTLLEGSRTVGRVPAERMDAAAMPDP 60
QY 61 DPDAAGKTPVTRASFSLDVACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIAPS 120
Db 61 DPDAAGKTPVTRASFSLDVACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIAPS 120
QY 121 ALVGTETVFTIGIGSEYEALPQATASAEIDAHGGLCTMPSVGARISYALGRPCVA 180
Db 121 ALVGTETVFTIGIGSEYEALPQATASAEIDAHGGLCTMPSVGARISYALGRPCVA 180
QY 181 VDTAYSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLWLSKTPALARDGCKAFS 240
Db 181 VDTAYSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLWLSKTPALARDGCKAFS 240
QY 241 AADGFGREGCAVVLKRLSGARADGDRILAVIRGSAINHGAASSGLTVENGSSQEIYL 300
Db 241 AADGFGREGCAVVLKRLSGARADGDRILAVIRGSAINHGAASSGLTVENGSSQEIYL 300
QY 301 KRALADAGCAASVGYVEAHGTTGLDPIEIQALNAVYGLGRDVAATPLLIGSVKXNLGH 360
Db 301 KRALADAGCAASVGYVEAHGTTGLDPIEIQALNAVYGLGRDVAATPLLIGSVKXNLGH 360
QY 361 PEYASGITGLKVLSTLOHGOIPALHQAOLNPRISWDDLLTVTRATPMDMTPRRA 420
Db 361 PEYASGITGLKVLSTLOHGOIPALHQAOLNPRISWDDLLTVTRATPMDMTPRRA 420
QY 421 GVSSFGMSGTAHVLEBPAPATCTPPAPERPAELLVISARTASALDQAARLRDHELETY 480
Db 421 GVSSFGMSGTAHVLEBPAPATCTPPAPERPAELLVISARTASALDQAARLRDHELETY 480
QY 481 PSQCLGDVAFSLATTRSAMEHRLAVALTSRGLRAALDAAAOQGTSPGAVSIAOSSRGK 540
Db 481 PSQCLGDVAFSLATTRSAMEHRLAVALTSRGLRAALDAAAOQGTSPGAVSIAOSSRGK 540
QY 541 LAFIPTGGAGTTLGMRGLYDVWSAFREAPDLCVRLFQOEIDRLREVMAMEPASPVDAL 600
Db 541 LAFIPTGGAGTTLGMRGLYDVWSAFREAPDLCVRLFQOEIDRLREVMAMEPASPVDAL 600
QY 601 LDQTAFTQPALFTEFEYALALMRSGVEPELVAGHSIGELVAAQVAGFSEDAVFLVAA 660
Db 601 LDQTAFTQPALFTEFEYALALMRSGVEPELVAGHSIGELVAAQVAGFSEDAVFLVAA 660
QY 661 RGRIMQALPAGGAMVSIAPADVAVAAPHAASVSIANAAPDQVVIAGAGQPVHATAA 720
Db 661 RGRIMQALPAGGAMVSIAPADVAVAAPHAASVSIANAAPDQVVIAGAGQPVHATAA 720
QY 721 AMARAGARTKALHSHAFSPMLAPMLEAFGRVAVESVYRRPSITLVNLSGKACTDVS 780
Db 721 AMARAGARTKALHSHAFSPMLAPMLEAFGRVAVESVYRRPSITLVNLSGKACTDVS 780
QY 781 SPQVWVRHAREVAFADGVKALHAAAGCTFVEVGKSTLLGLVPACMPDAPALLASSRA 840
Db 781 SPQVWVRHAREVAFADGVKALHAAAGCTFVEVGKSTLLGLVPACMPDAPALLASSRA 840
QY 841 GRDEPATVLEALGGLMAVGLSVNAGLPSSGRRVPLFTYPMQRRYVIDTKADDAAGD 900
Db 841 GRDEPATVLEALGGLMAVGLSVNAGLPSSGRRVPLFTYPMQRRYVIDTKADDAAGD 900
QY 901 RRAAGAGDEVEEGGAVRGDRRSARLDHPPESGRREKVAAGRPRLIIDEVGLD 960
Db 901 RRAAGAGDEVEEGGAVRGDRRSARLDHPPESGRREKVAAGRPRLIIDEVGLD 960
QY 961 LVLRTERRABGLSEVEIAVDAAGLSPNDVOLALGMVDDLPGRKPNPLLOGECAGTIV 1020
Db 961 LVLRTERRABGLSEVEIAVDAAGLSPNDVOLALGMVDDLPGRKPNPLLOGECAGTIV 1020
QY 1021 AVGSGVGLVVGQPVIALSAGAFATHTTSALVLPDQALSAIEAAMPVAYLTAWYAL 1080
Db 1021 AVGSGVGLVVGQPVIALSAGAFATHTTSALVLPDQALSAIEAAMPVAYLTAWYAL 1080

```

```

Db 1021 AVGSGVGLVVGQPVIALSAGAFATHTTSALVLPDQALSAIEAAMPVAYLTAWYAL 1080
QY 1081 DRIARLOGERVULHAATGGVGLAAVOACHVGAEVHATAGTPEKRAYLESIGRYVSDS 1140
Db 1081 DRIARLOGERVULHAATGGVGLAAVOACHVGAEVHATAGTPEKRAYLESIGRYVSDS 1140
QY 1141 RSDRFVADVRAMTGEGVDVVLNSLSELIDKSFNLLRSHGRPYELGRDCYADNOQLGR 1200
Db 1141 RSDRFVADVRAMTGEGVDVVLNSLSELIDKSFNLLRSHGRPYELGRDCYADNOQLGR 1200
QY 1201 PFLRNLFSFLVDLFGMMLERPARVALLBEELGLIAGVFPFPPIATPIARVADAFRSM 1260
Db 1201 PFLRNLFSFLVDLFGMMLERPARVALLBEELGLIAGVFPFPPIATPIARVADAFRSM 1260
QY 1261 AQAHLGLVLTLDPEVQIRIPTHAGAPSTGSRDLDRLASAPARAALAEFLRTQ 1320
Db 1261 AQAHLGLVLTLDPEVQIRIPTHAGAPSTGSRDLDRLASAPARAALAEFLRTQ 1320
QY 1321 VSOVLRTPEIKVGAELFTRIGMDSLMAVELRNRIEASLKXKLTSTPSPNIALLAQN 1380
Db 1321 VSOVLRTPEIKVGAELFTRIGMDSLMAVELRNRIEASLKXKLTSTPSPNIALLAQN 1380
QY 1381 LLDALATRLSLERVAENLRAGVONDRVSGADODWEITIAL 1421
Db 1381 LLDALATRLSLERVAENLRAGVONDRVSGADODWEITIAL 1421

```

```

RESULT 8
US-09-335-409-7
; Sequence 7, Application US/09335409.
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Jelvian
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-335-409-7

Query Match      57.4%; Score 4137; DB 3; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8.5e-287;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

QY 6 IERAADPIAIVGASCRLLPGGVIDLGGFWTLLEGSRTVGRVPAERMDAAAMPDPDPAP 65
Db 26 LEQERTERPIAIVGICRFPFGADTRPEAFWELLDGRDAV--QPIDRKALVGNPSEVP 83
QY 66 GKTPTVTRASFSLD-VACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIAPSALVG 124
Db 84 -----RMAGLTTEAVADGFDAFFGTSPREARSIDPQORLLLEVWEGLEDAGIAPQSIDG 138
QY 125 TENGVPFIGIGSEYEALPQATASAEIDAHGGLCTMPSVGAGRISYALGRPCVAYDTA 184
Db 139 SRTGVFLGAGSDVSHTVAAQORRE-EQDAVDITENTLISVAAGRISYTLGLQGCLTVDTA 197
QY 185 YSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLWLSKTPALARDGCKAFSAEAD 244
Db 198 CSSSLVAVHLACRSLRARBEDIALAGVNMLSKTMIMIGRIQALPDPGCRFPDASAN 257
QY 245 GFGREGCAVVLKRLSGARADGDRILAVIRGSAINHGAASSGLTVENGSSQEIYLRAL 304
Db 258 GFVRGEGCGMVLKRLSDAQRHGDRIWALIRGSAMNODGSRSTGIMANVLVLAQALREAL 317

```

QY 305 ADAGCAASSVGYVEAHGCTTIGDPIETIOALNAVYGLGDVATPPLLIGSVKTNLGHPEYA 364  
Db 318 OSARVDAGIIGVEETHGCTSLGDPIEVAPALVIGPARADSSRCVLGAVKTNLHLLEBA 377  
QY 365 SGTIGLKVLSLQGOQPAHLHAQALNPRISMGDLRLVTARATPMPMTPRAGVSS 424  
Db 378 AGVAGLIRKALHLHLLPRNHPHTLNPRIRIEGALLALATEPVPMPAGPRRAGVSA 437  
QY 425 FGMSTNAHVLEAPPAATCTPPAPERPAELLVLSARTASALDAQAARLDHLETTYPSQC 484  
Db 438 FLSGTNVHVLEAPATVATPATPCRSALLVLSKSAALDAQAARLSAHIAAYPEQG 497  
QY 445 LGDVAFSLATTSMAEHRLAANAATSEGLRALDLDAAGOTSPGAVRSIADSSRKLAPL 544  
Db 498 LGDVAFSLVSTSPMEHRLAANAATSEGLRALDLDAAGOTSPGAVRSIADSSRKLAPL 557  
QY 545 FTGGAQTLGMRGGLYDVMSAFREAFDLCLVLFNOELDRPLREVMMAEPASVDALLDOT 604  
Db 558 FAGGQAQVPGMGRGLMEAMPAREFTFDRCVTLFDEELHOPLCBVMAEBSRSSLLDOT 617  
QY 605 AFTQPALFTFEVYALALMWSWGBELVAGHSIGELVAACVAGVSLBDVFLVARGEL 664  
Db 618 AFTQPALFTFEVYALALMWSWGBELVAGHSIGELVAACVAGVSLBDVFLVARGEL 677  
QY 665 MOALPAGGAMVSIAPADVAANAAPHAASVSIANAAPDOVYIAGAGPVAIAAAMA 724  
Db 678 MOALPAGGAMVSIAPADVAANAAPHAASVSIANAAPDOVYIAGAGPVAIAAAMA 737  
QY 725 RGARTKALVSHAFSPPLMAFMLFAFGRVAESVSRPSIVLVSNLGKACTDEVSPPQY 784  
Db 738 RGARTKALVSHAFSPPLMAFMLFAFGRVAESVSRPSIVLVSNLGKACTDEVSPPQY 797  
QY 765 WVRHAREVVRFPADGVKALHAAGAGTFVEVGPSTLGLVAPACPDARPALASSPAGDE 844  
Db 768 WVRHAREVVRFPADGVKALHAAGAGTFVEVGPSTLGLVAPACPDARPALASSPAGDE 857  
QY 845 PATVLEALGELMAVGVLSVMAGLFPGSGGRVPLPYPMQREKRYM----DTKADDAAR-- 898  
Db 858 AASALEALGELMAVGVLSVMAGLFPGSGGRVPLPYPMQREKRYM----DTKADDAAR-- 917  
QY 899 ----- 898  
Db 918 AGGHPILGEVFSVTHAGRLMETTLDRKRLPWLGBHRAQSVFPAGVLEMASSGAE 977  
QY 899 -----GDRRAP----- 904  
Db 978 ILGDGPQVTVLLETTLFAGDTAVPVQVVTTEERPGRLRFQVASREGERRAPRIHA 1037  
QY 905 ----- 904  
Db 1038 RGVLRIRIGVETPARSNLALRLARLHAAPVAPAAIYGALAEMLQYQPALRGIAELMRGEG 1097  
QY 905 ----- 904  
Db 1098 BALGRVRLPEAASATAYQLHPVLLDACVQMTVGAFAORDDEATPMAPEVGVSVRLFQSP 1157  
QY 905 ----- 904  
Db 1158 GELMCHARVSDGOASSSRMSADPELMDGTGAVVAEISRLVVERLASGVRRRADDWPLE 1217  
QY 905 -----GAGH----- 908  
Db 1218 LDMEPALGPKITTAGRWLLLEGGGLGRLCSALKAAGHVVAAGDDTSTAQRALIA 1277  
QY 909 ----- 908  
Db 1278 NAFDQAPTAVVVHSSLDGGGLGRLGAGALDAPRSPVDADALLESALMRGCSVSL 1337  
QY 909 ----- 908  
Db 1338 VOALVGMDLRNAPLMLTRGAQAAAAAGDVSVVQAPLLGLGRTIALHEALRCSVDLP 1397

QY 909 -----DEVEGAVRGDRRSARLDHPPEPSGREKVEAAGDRPFRL 952  
Db 1398 AEPGEADALLAELLADDAEERVLRGDRVLVAALVHLEPRAORKEKVEPPGDRPFRL 1457  
QY 953 DEPGVLDLVLVTERARAPGLGEVEIAVDAAGLSFNVDQALGAWPDDLPCKPNNPLL 1012  
Db 1458 DEPGVLDLVLVTERARAPGLGEVEIAVDAAGLSFNVDQALGAWPDDLPCKPNNPLL 1517  
QY 1013 GECAGRIYAVEGVNGVLVGPVIALSAGAFATHTTSALVLPPOALSIAEAAAMVA 1072  
Db 1518 SECAGRIYAVEGVNGVLVGPVIALSAGAFATHTTSALVLPPOALSIAEAAAMVA 1577  
QY 1073 YLTAMVADRIARLOPGERVLIHAATGVGAIAVQMAOVGAEBVATAGTEPKAYLESL 1132  
Db 1578 YLTAMVADRIARLOPGERVLIHAATGVGAIAVQMAOVGAEBVATAGTEPKAYLESL 1637  
QY 1133 GVRVYSDRSRFPVADVAPAMTGECEVDVNLNSGELIDKSFNLRSRSHRFEVLEKRDY 1192  
Db 1638 GVRVYSDRSRFPVADVAPAMTGECEVDVNLNSGELIDKSFNLRSRSHRFEVLEKRDY 1697  
QY 1193 ADNOLGRLPPLRNISFSLVDLRGMMLERPAVRLLEELGLIAGV----- 1239  
Db 1698 ADNOLGRLPPLRNISFSLVDLRGMMLERPAVRLLEELGLIAGV----- 1286  
QY 1240 FTTPPIATLPARVADAFRSMAOQHLGKLVTLTGDPPEVOIRIPHA----- 1289  
Db 1758 FTTPPIATLPARVADAFRSMAOQHLGKLVTLTGDPPEVOIRIPHA----- 1817  
QY 1287 -----GAG----- 1289  
Db 1818 GGLGLGLRVAGVWMLAERGAQGLVLVGRSAGASABORAAVALAENHARVAKADVADS 1877  
QY 1290 ----- 1289  
Db 1878 QIERVAREVTASGMPRLRGVHAAGLVDDGLMOQTPAFRTVMGPKVQAGALHHTLREA 1937  
QY 1290 ----- 1289  
Db 1938 PLSFVVLVYASAGLFPGSGGQNYAANAFLDALSHHRAQGLPALSIDMGFTVEGMAVA 1997  
QY 1290 -----BST----- 1292  
Db 1998 QENRGARQISRMGKITPDEGLSALARLLEGDRVQTVIPIPPQWVEFYATAASRLS 2057  
QY 1293 -----GDRPLRLASAPAAALAEFLRQVQVLTPEIKYGAELF 1338  
Db 2058 RLVTQRAVADRTAGDRDLLEQLASAEPSARAAGLQDVVRQVSHVRLPEDKIEVDAPL 2117  
QY 1339 TRLGMSIMAVELENNRIEASLKLXSTPFLSTSPNIALAONL-DALATAL 1389  
Db 2118 SSMGMSLSMSELNRRIFALAGVAPALGWTYPTVAAITRWLDDALVRL 2169  
  
RESULT 9  
US-09-568-102-7  
; Sequence 7, Application US/09568102  
; Patent No. 6346404  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Jettan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568.102  
; PRIOR FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335.409  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 7

```

; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-7

```

Query Match	57.4%;	Score 4137;	DB 4;	Length 2439;
Best Local Similarity	43.4%;	Pred. No. 8.5e-297;		
Matches 933;	Conservative 135;	Mismatches 308;	Indels 776;	Gaps 13

Qy	6	IEEAAADPLAIWASCRLEGGVVIDLSGFMTLLGGSDTVGVAPREMDMAAMFPDDAP	65
Dp	26	LEBERKEPFLAIWIGICRFFPGADTBFAFHELLDSRDV--QPLDRKALVGNPSEVP	83
Qy	66	GKTPVTRASFSD-VACPDASFEGISPREALRMDPAHLLLEVCMEALENNAIPSAVLG	124
Dp	84	-----RMAGLITBAVDGFDAAFFGISPRARSLDPQQLLLEVTWEGIEDAGINPGLDG	138
Qy	125	TEVCVITIGIPSEYEALPQATASAEIDAHGGLGTPSVGAGRIISYALGRFCVAVDTA	184
Dp	139	SRGVFLGACSSDYSHTVAQORRE-EODAYDITGNTLSVAAGRIISYTLGIQGPCLTVDTA	197
Qy	185	YSSSLVAVHACOSRSGECSTALAGVMSIPSTLWMLSKTALARDGCKRFSAD	244
Dp	198	CSSSLVAHACRSRLARASDIALAGVWMLSSXTMTLMGRIGQLSPDGCRRFDSAN	257
Qy	245	GFGRGEGCAVVYLKRLSGARADGDIRLLVIRGSAIINHDAAGSGLTPVPGSOEIVLRAL	304
Dp	258	GFVRGGCGAMVYLKRLSDAQHRGDRIWMLIRSANNOGRSTGLMAPRVLLAQEALREAL	317
Qy	305	ADAGCAASSVGEYEAHGTTTGADPIETIOALNAVYGLGRDVATPLLIGSVKTNIGHPEYA	364
Dp	318	QSARVDAGAIQVETETGTSTLGDPIVEEALPAVLGPARADAGSCVIGAVKTNIGHGEA	377
Qy	365	SGITLKLKVVLSIQHOQIPAHILHAQALNPRIWMGDLRTVIRARTPMDMTPRBAVSS	424
Dp	378	AGVAGLIKALAHHEHLLPRNLHFTTLMNRIETSTALATTEPYPMWRAPRPFASVA	437
Qy	425	FGNSGTNAHVLEEARPAATCTPAPERPAELLVLSARTASALDAQAARLDHETTESQC	484
Dp	438	FGLSGTNAVHVEEARATVLAATPATGRSALLVLSAKSAALLDAQAARLSHIAAYEQG	497
Qy	485	LGVASLATTSSANEHLAAVAATSREGIRAAALDAAGQOTSPGAVNSIADSGKIAFL	544
Dp	498	LGVASLSTRSPMEHRLAVALATSRREALRSLBEVAAQOTPAGAAGRASSPGKIAFL	557
Qy	545	FTGGGQTLGMSGGLVDVSAFREA.PDLCVRLFNGEELRPLREVMMAEPASVDAALLDOT	604
Dp	558	FACGGQVPMGKGLMEANPAFRETFDRCVTLFDELLHQPLCEVMMAPGSGRSRLDQT	617
Qy	605	AFTQPALFTFEYALALMRMGSVEBELVAGHSIGELVAACVAGVSELEDAFLVAARGRL	664
Dp	618	ALFQPALFLYELALALPRMGSVEBELVAGHSGELVAACVAGVSELEDAFLVVAARGRL	677
Qy	665	MQALPAGAMVSIIEAEADVAALVA.PHAASVSIIAVNA.PDQVITAGAQPYHAIAMAA	724
Dp	678	MQALPAGAMVSIIEAEADVAALVA.PHAALVSIIAVNPBEQVITAGAEKFFQIILAAFAA	737
Qy	725	RGARTALHVSIAHFHPLMA.PMLLEAPGRVAESVSTRRSYLVNLSGKACTDEVSSPGY	784
Dp	738	RGARTPLHVSIAHFHPLMD.PMLEAPRRTESVTRYRPSIALVNSLGSKPCCTDEVSPGY	797
Qy	785	WVYHAREVVRFPADGVYALHAAGAGFVEVGPSSTLLGLVPACMPDAPALLAASRARDE	844
Dp	798	WVRHAREVVRFPADGVYALHAAGAGFVEVGPRPTLLGLVPRACPDARVLLPASRARDDE	857
Qy	845	PATVLEALGLVAVAGLVSMAGL.PPSGGRRVPLPTYMQORERYWI-----DTKADDAAR--	898
Dp	858	AASALEALGFVVVGSGVTWVGSPFSGGRVPLPTYMQORERYWIEAPVDEADGTGRAR	911
Qy	899	-----898	
	918	AGGHPLLGGEVSVSTHAGLRLWETLLDRKRLPMIGEGHRAQGEVVPFAGYLINALSSGAE	977

[illegible]

Db 2058 RLVTTORAVADRTAGDRDLLEQLASBPBARAGLQDVVRVOVSHVLRIRPEKXIVDADL 2117  
Qy 1339 TRLGWDSLMAVELRNRIEASLKLTSTFLSTSPNIALIAQNL-DALATAT 1389  
Db 2118 SSMGDSIMSLRNRIRIEALGVAAAPALGWTYPTVAATRWLIDLADLVRL 2169

RESULT 10  
US-09-567-969-7  
Sequence 7, Application US/09567969  
Patent No. 6355457  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligot, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/567,969  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 7  
LENGTH: 2439  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-567-969-7

Query Match 57.4%; Score 4137; DB 4; Length 2439;  
Best Local Similarity 43.4%; Pred. No. 8.5e-297;  
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

Qy 6 IERAEDPPIAVGASCRIPGIVIDLSGFWTLEGSRDYGRVPRKWDAAAFDDPDPA 65  
Db 26 LEQERTPIAVIGICRPFPGADTPBEAFWELIDSGDAV--QPLDRMLVGVPHEEVP 83  
Qy 66 GKTPTRAFSLSD-VACFDSPFGISPREALRMDPAHRLLEVCWEALENAIAPSALVG 124  
Db 84 ----RMAGLITFADVDFPAATFGTSPPRARSLDPOQRLLLEVTEGLEDACIAQSLDG 138  
Qy 125 TETGVFIGSESEYEALEFQATASARIDAHGGLGTMPYVAGRISYALGLRCPVAVDTA 184  
Db 139 SRTGVFLGCSSDYGHVTAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLQGPLTVDTA 197  
Qy 185 YSSSLVAVHLACQSLRSRGCSTALAGVSLMSPSTIVLSTKRLAARPKAKASLEAD 244  
Db 198 CSSSLVAHLACRSIRARSDIALAGVWMLSSKTMILGRIOALSPDGHCRTPDASAN 257  
Qy 245 GFGREGGAVVVLKSLGSRADGDRILAVIRGSAINHDGASSGLTVPNSSQOEVILKRL 304  
Db 258 GFVREGGCMVVLKRLSDQQRHGRHRIWALIRGSANQDGRSTGLMAFVNLAGEALLREXL 317  
Qy 305 ADAGCAASSVGYEVAHGTTTLGDPRIEIOALNAVYGLGRDVAFTPLIGSVKTNLGHPEYA 364  
Db 318 QSAVDAGAGIGVEYTHGTSTLGDPIEVEALRAVLPARADGSRCLGAVKTNLGHLEGA 377  
Qy 365 SGIITLLKVLVSLQGOIPAHILAAQALNPRISMGDLRLTVTAARFPMWMTPRRAGVSS 424  
Db 378 AGVAGLITKAALHHELLPRNLHFTLNRIRIEGTALALATEPVWPAPAGRPRAVISA 437  
Qy 425 FGMSTNAHVLEAPAACTPPAPBRPAELLVLSTARTASALDAQAARLDLETPSQ 484  
Db 438 FGLSGTNHVLEAPATVLAATPGRSALLVLAKSAALDAQAARLSAHIAHSPBG 497  
Qy 485 LGDVAPSLATTSAMEHRLAVALTSRBEGLRALDAQAQOTSFGAVRSIADSSRGKLAFL 544  
Db 498 LGDVAFSLVSTRSPMEHRLAVALTSREALRLSLEVAQAQOTPAGAAAGRAASPGKLAFL 557

Qy 545 FTGGAQTLLMGNGRLYDVMSAFREAFDLCVRLFNOELDRPLREVMMAEPASVADALLDOT 604  
Db 558 FAGGAQVPGMGKGLMAMPAFRETFRDRCVTLFDBELHQLPCEVMMAEPGSSRSLLDOT 617  
Qy 605 AFTQPAFTFEYALAAALRSWGEVEBELVAGHSIGELVAACVAGVSELDAYFLVARGRL 664  
Db 618 AFTQPAFLALEYALAAALFRSWGVEBELVAGHSIGELVAACVAGVSELDAYFLVARGRL 677  
Qy 665 MQALPAGAMVSIAPADVAANAAPHAAVSIANAANPDQVVIAGAQOPHATIAAAMA 724  
Db 678 MQALPAGAMVSIAPADVAANAAPHAAVSIANAANPDQVVIAGAQOPHATIAAAMA 737  
Qy 725 RGARTKALVSHAFSPMAAPMLBAFGRVAVSVSYRPSIYLVNSLSKACTDEVSPPGY 784  
Db 738 RGARTKPLVSHAFSPMLDPMLEAFRVTSTYTRRSIALVNSLSKPCPTDEVSABGY 797  
Qy 785 WVRHAREVVRPADGVKALHAAAGCTFVEVGKSTLLGLVPACMPDARALLASSRAGDE 844  
Db 798 WVRHAREVVRPADGVKALHAAAGCTFVEVGKPTLLGLVPACLPDARVLLPASRAGDE 857  
Qy 845 PATVLEALGLMANVGLVSNAGLPSPGRRVPLPTYPQORRYMT----DTKADDAAR-- 898  
Db 858 AASALEALGFMVVGSGVTWVGSPGGRVPLPTYQORRYMTLEADVDEADGTGRAR 917  
Qy 899 ----- 898  
Db 918 AGGHPILGEVSVSTHAGRLMETTLDRKRLPMLGHRHQAQGVVFPAGCYLEMALSSGAE 977  
Qy 899 -----GRRAP----- 904  
Db 978 ILGDPIQVTVLIETLTFAGDPVAPVQVTTERRPGLRFOVASREPGERRAPFRJHA 1037  
Qy 905 ----- 904  
Db 1038 RGVLRIRIGRVETPARSNLAALRRLHAAVPAALYALGALEMGLQYGPALRGLAELMRGEG 1097  
Qy 905 ----- 904  
Db 1098 EALGRVRLPEAAGSATVQLPVLIDACVQMLVGAFAIRDENTPAPVSVSVLRFORSP 1157  
Qy 905 ----- 904  
Db 1158 GELMCHARVVDGQOASRMSADFELMDGTGVAVAEISRLVVERLASGVRRDADWFLE 1217  
Qy 905 -----GAGH----- 908  
Db 1218 LDWEPALGPKITAGRWLLLBGGGLGRSLCSALKAGHVVAHAAGDSTAGKALLA 1277  
Qy 909 ----- 908  
Db 1278 NAFDQAFNAVHLSLSDGGQLGFGLAGAQAALDAPRSPVDADALBESALRGCDVSL 1337  
Qy 909 ----- 908  
Db 1338 VQALVGMDLRANAPRLMLITRGAQAAAAGVSVQAPLLGLGRTIALBEAHELRCSIVDDP 1397  
Qy 909 -----DEVEEGAVRGDRRSALDHPSPSSGREGREKTEAAGDPRFREI 952  
Db 1398 ABEPEGEDALLAEILLADDAEEVVALRGDDRVARIVHMLPAQOREKVEPEGDPRFREI 1457  
Qy 953 DEPGVLDLVLVTRERRAPGLGEVEIADVADAGLSFNDVQLALGWPPDDLPCKPMPPLLG 1012  
Db 1458 DEPGALDVLVIRACRRAPRGVEVEISVEAAGLDSIDQLALGVAPNDLPBEIEIPLVIG 1517  
Qy 1013 GECAGRIVAVGEGVGLVVGQPVIALSAGAFATHTTSALVLPPOALSAIEAAMPVA 1072  
Db 1518 SECAGRIVAVGEGVGLVVGQPVIALAAGVATHTTTSATVLPBPLCLISATEAAMPVA 1577  
Qy 1073 YLTAMVADLRARLQOPGERVLIHAATGVGLAAVQMAQVGAHYHATGTGTEKAYVLESL 1132  
Db 1578 YLTAMVADLRKVAHLOQGERVLIHAEAGGVGLCAVRMAQVGAHYATADTEPNRKYVLESL 1637  
Qy 1133 GVRVYSDSRDFADVAVAMTGGEGVDVVLNSLSEGLIDKSFNLRLSHRGFEVLEGRDCY 1192



```

Db      1638 GVRVVSRSRGRFVTDVAAWTDGEGVDVLDLSGERIDKSLMVLKAGRLVKGRDDC 1697
Qy      1193 ADNOLGRLPRLRNISFSLVDLRGWWLBERPARVRLLEELGLIAGV----- 1239
Db      1698 ADTPGLPRLRNISFSGVDLRGWWLDDPARLRALDELFGVLAAGALSPLGSLRGGS 1757
Qy      1240 FTTPPIATLPARVADAFRSMAOQHLGKLVLTGLDPEVOQIRIPTHA----- 1286
Db      1258 LTPPEVETPISRAAEFRMAOQHLGKLVLTLDPEVRIRAPRESSVAARADOTYLV 1817
Qy      1287 -----GAG----- 1289
Db      1818 GGLGLGLRVAGWLAERGAQQLVLRGSGAASQRAAVALAAGARVAKADVADRS 1877
Qy      1290 ----- 1289
Db      1878 QIERVLEVTASGMPRLRGVHAAGLVDDGLMOQTAPAFRTVMGPKVQALHHTLREA 1937
Qy      1290 ----- 1289
Db      1938 PLSPFVLVYASAGLEFGSPQGNVAAANAFDLALSHHRAQGLPALSIDMKFTVGMVA 1997
Qy      1290 -----PST----- 1292
Db      1998 QENRGARQISRGMGITPDEGLSALARLLEGDRVQGVIPITPRQWVEFYATAASRLS 2057
Qy      1293 -----GDRDLRLASAPARAPAALEALRTQVSGVLTPEIKVCAELF 1338
Db      2058 RLVTTPRAVDRTAGDRDLLEQLASAEPSARAGLLQDVAVQVSHVRLPEDKLEVDAPL 2117
Qy      1339 TRLGMDSLMAVELNRIRIASLKLKLTFTLSTSPNIALAQNLL-DALATAL 1389
Db      2118 SSMGMDSLMSLELRIRIALGVAPALGWTYFVAATRWLDDALVRL 2169

RESULT 11
US-09-568-480-7
; Sequence 7, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt. Devn
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-7

```

```

Query Match      57.4%; Score 4137; DB 4; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8,5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

```

```

Qy      6 TERAEDDIAIVGASCRLLPGVITDLSGFWTLLGSRDVTGVRPAERMDAAAFDDPDAP 65
Db      26 LEOETEDIAIVGICRPPGADTPEAFWELLDSDRDV--QPLDRRAVLGVHSESEVP 83
Qy      66 GKTPTTASPLSD-VACEDASFFGISPREALRMDPAHLLLEVCHEALEMAIASALVG 124
Db      84 -----RWAGLLTEAVDGFDAAFGTSPPARSLDPOQRLLLEVTEGLEDAIGIAQSLDG 138

```

```

Qy      125 TETGVFIIGISPEYEALPQATASAEIDAHGGIGTMSVGAGRISVYLGRLGPCVAVDTA 184
Db      139 SRTGVFLGACSDSYSHYTAQORRE-EQDAYDITGNTLSVAAGRLSTYLGQPCCLTVDTA 197
Qy      185 YSSSLVAVHLACQSLRSGEGSTALAGVSIMLSTLVMLSKTLARLDRGCKAFSAEAD 244
Db      198 CSSSLVAIHACRSLRREEDLALAGVNMILSKTKIMIGRIQALSPDGCRIFDASAN 257
Qy      245 GFRGECQAVVVLKRLSGARADGRILAVIRGASINHDGASSGLTPNGSQEIVLKRAL 304
Db      258 GFVARGCGMVLKRLSDAQHGRDRIWALIRGSMNDGSRGTGMARVLAQEAELRLAL 317
Qy      305 ADAGCAASVGYEBAHGTGTLGDPRIQALNAVYGLGRDVAIPLIGSVKTNLGHPEYA 364
Db      318 QSAVDAAGALGYVETHGTGTSIGDPRIEVALRAVLGPARADGSRVCVGAVTNIGHLEGA 377
Qy      365 SGITGLLKVVLSTLOHQIPAHILAAQALNPRISWQDLTLVTYRATPPMDMTPRRAGSS 424
Db      378 AGVAGLIKAALALHHELIPLHLFHTLNPRIIRIGTALALATEPEVPPRAGRPPAGVSA 437
Qy      425 FGMGTNAHVVLJEAPATCTPPAPERPAELVLVSARTASALDQAARLRDHLSTYPSQC 484
Db      438 FGLSGTNHVVLJEAPATVLAIPATPGSABELLVTSASAAALDQAARLSHIAIAYPEQG 497
Qy      485 LGDYAFSLATTRSAMEHRLAAVATSREGRLAALDAAAQGTSPGAVSIAOSSRGKLAFL 544
Db      498 LGDYAFSLVSTRSPMEHRLAAVATSREALRSALVAAQGTGPAGAGRAASPGKLAFL 557
Qy      545 FTGGCAQTLGMRGLYDWSAFREAFDLCYRLFNOELDRPLREVMMAEPASVDAALLDOT 604
Db      558 FAGCAQVPMGMRGLMEAMPARETFDRCVTLFPRELHQPCVEMVMEPGSSRSLLDOT 617
Qy      605 AFTQPALFETFEYLAALMRSWGVPELVAGHSIGELVAAVCAVGFSLVEDAFVLAARGRL 664
Db      618 AFTQPALFALEYLAALAFRSWGVPELVAGHSIGELVAAVCAVGFSLVEDAVRLVYARGRL 677
Qy      665 MQALPAGAMVSIAPADVAVAAPHAASVSIANAAPDQVITAGAGQPVHAIAPAAVA 724
Db      678 MQALPAGAMVSIAPADVAVAAPHAALVSIANAANGPEQVITAGAEKFOQIAAPAA 737
Qy      725 RGARTKALHSHASHSLPMAEMLEAFGRVAESVSRPSIVLVNLSGKACTDEVSSGXY 784
Db      738 RGARTKPLHSHASHSLPMPMLEAFRRVYESVYTRPSIALVSNLSGKPCCTDEVSAGXY 797
Qy      785 WVRHAREVRPADQVKLLHAAGACTPVEVGKSTLLGLVPCMPDAPRALASSRAGDE 844
Db      798 WVRHAREVRPADQVKLLHAAGAGLFEVVGKPTLLGLVPCMPDAPRALASSRAGDE 857
Qy      845 PATVLEALGILMAVGLVSMAGLFPSSGRRVPLPTYPQORERYWI---DTKADDAAR-- 898
Db      858 AASALEALGFMVVGSVTWSGVFPSSGRRVPLPTYPQORERYWI---DTKADDAAR-- 917
Qy      899 ----- 898
Db      918 AGHPHLLGEVPSVTHAGLRIMETTLDRKRLPWLGEHRAQGEVFPAGYLEMALSSGAE 977
Qy      899 -----GDRAP----- 904
Db      978 ILDGPIQIVDVVLIETLTFRAGDTAVPVQVVTTEERGRRLFOVASREPGERRAPFRTHA 1037
Qy      905 ----- 904
Db      1098 RGVLRIGRVETPARSNLALRLARLHAAPPAALYGALEWGLQYPALRGLAELMRGEG 1097
Qy      905 ----- 904
Db      1098 EALGRVRLPEAAGATAYQOLHPVLLDACVQMIVGAFADRDPEATPAVGVSVRLFORSP 1157
Qy      905 ----- 904
Db      1158 GELMCHARVVSDDQOASSRMSADELMDGTGAVVAEISRLVVERLASGVRRRDDDWTFLE 1217
Qy      905 -----GAGH----- 908

```

```

Db      1218 LDMEPALGSPKITAGRWLLLEGGGLGRSLCSALKAAQHVVAAGDSTTAGKRALIA 1277
QY      909 -----
Db      1278 NAFDQAPTAVVHLSLDGGGQLGPGLAGAQLDAPRSPDVADALLESALMGCDVSLSL 1337
QY      909 -----
Db      1338 VQALVGMRLNAPRLMLTRGAQAAAAAGVSVVQAPLLGLGRTIALEHAEIRCSVDDLP 1397
QY      909 -----
Db      1398 AEPBEADALALAEADDAEEVALRGDRVLARLVHRPDQORKEKPEADRPFRLEI 1457
QY      953 DEPGVLDHVLVTERBARGLCEVEIAYDAAGLSFNDVOLALGMVDDLPKGNPPLIG 1012
Db      1458 DEPGALDQVLTRATGRAPGSEVEISVEAAGLDSIDQLALGVAPNLPGSEIPLVIG 1517
QY      1013 GECARIVAVGCVNGVLVVGQVIALSAGAFTHYTTSAALVLRPQALSAIEAAMPVA 1072
Db      1518 SECARIVAVGCVNGVLVVGQVIALSAGAFTHYTTSAALVLRPQALSAIEAAMPVA 1577
QY      1073 YLTAVYALDRILARLPGERVLIHAAATGGVGLAAVQMAOHVGAHVHATAGTPEKRAYLESL 1132
Db      1578 YLTAVYALDRILARLPGERVLIHAAATGGVGLAAVQMAOHVGAHVHATAGTPEKRAYLESL 1637
QY      1133 GVRVYSDSRDRFVADVRAWTGSEGVVVVNSLSEGLIDKSFNLLRSHGRFVELGRDXY 1192
Db      1638 GVRVYSDSRGRFVADVRAWTGSEGVVVVNSLSEGLIDKSFNLLRSHGRFVELGRDXY 1697
QY      1193 ADNOLGRFPELNSFSLVLDLRGMMLERPARVRLLEELGLIAAGV----- 1239
Db      1698 ADTQGLRPLNLFNSFSDVLDLRGMMLERPARVRLLEELGLIAAGV----- 1757
QY      1240 FTTPPIATPLIARVADAFRMAOQHKLVLTLGDPVEYQIRIPTHA----- 1286
Db      1758 LTPPEVETFPISRAAEAFRMAOQHKLVLTLGDPVEYQIRIPTHA----- 1817
QY      1287 -----GAG----- 1289
Db      1818 GGLGGLGLRVAAGMLAERGAQQLVLVGRSAGAAEQRAVALEAHGARVYAKADVADR 1877
QY      1290 ----- 1289
Db      1878 QIERVLEVTASGMPLRGVVAAGLVDDGLMQQTPARFRYMGPRVQALHLHTLTREA 1937
QY      1290 ----- 1289
Db      1938 PLSPFVLVYASAGLFGSPQGVYAAANAFDLASHHRRAGQLPALSIDMGFTVGMAYA 1997
QY      1290 -----PST----- 1292
Db      1998 QENRGARQISRGMRGITPDEGLSALARLEGGDRVQGVPIPRQWVEFYRPATASRLS 2057
QY      1293 -----GDRLLDLRLASAAPARAAALEAFRTQVSQVLRPEELKVGAEALF 1338
Db      2058 RLVTTQRAVADRTAGRDLLLEQLASAEFARAGLLQDVVRVQVSHVLRPEEKIEVDAL 2117
QY      1339 TRLGDSLMAVEALRNRIEASLKLKLTSTPLSTSPNIALLAONLL-DALATAL 1389
Db      2118 SSMGDSLSLMSLELRRIEALGVAAAPALGMVYPTVAATLRWLDDALVVR 2169

```

```

; APPLICANT: Cyr. Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568, 486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 7
; LENGTH: 2439
; TYPE: PR
; ORGANISM: Sorangium cellulosum
; US-09-568-486-7

Query Match      57.4%; Score 4137; DB 4; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8.5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

QY      6 IERAAEDPIAIVGASCLRGVYIDLSGFWTLLEGSRTDVGVPAREMDAAMPDPDPAP 65
Db      26 LEOETEBIAIVIGICRPPGADTPEAFWEMLDGRDAV--QPIDRWALGVHSEVP 83
QY      66 GKTPTTRASFUSD-VACPDASFPGISPREALRMDPARHLLEVCMEALENAIAPALVG 124
Db      84 -----RWGLLLEAVDGDPAFFGTSPREASLDPOQLLEVEWEGLEDGIAPQSDG 138
QY      125 TETGVFIGISPEYEAALPQATASAEIDAHGGLGTPSVGAGISYALGLRQPCVAVDTA 184
Db      139 SRTGVFIGACSSDSVHTVAQQRRE-EDQAYDITGNTLSVAAGRLSYTLGLQPCLTVDTA 197
QY      185 YSSSLVAVHLCOSIRSECECTALAGVSLMSTLWLSKTLALARDGCKAFSAEAD 244
Db      198 CSSSLVAVHLCOSIRSECECTALAGVSLMSTLWLSKTLALARDGCKAFSAEAD 257
QY      245 GFGREGCAVVVLRISGARADGRIILAVIRGSAIINHGAAGSLTPVNGSOEIVLRAL 304
Db      258 GFRBEGCGMVVLRKLSDAQRHGRIMALINGANODGRGTGLMAPVVLQOELALREAL 317
QY      305 ADAGCAASVGVYEAHGTGTLGDPTEIOLANAYGLRDVATPLLIGSVKTNLGHPEYA 364
Db      318 OSARVDAIGAVETHEGTGTLGDPTEIOLANAYGLRDVATPLLIGSVKTNLGHPEYA 377
QY      365 SGITGLKVLVSLQHQIPALHHAOLNPRISKWDLRITVRATRPWDMNTPPRAGVSS 424
Db      378 AGVAGLIRKALALHHELIIRNLHPTTNIRIRIEGTALALATEPVMPRAGRPRAGVSA 437
QY      425 FGMSGTNAHVLEAPAACTPPAPERPAELIVLSARTASALDAQAARLDHLETPSPQC 484
Db      438 FGLSGTNHVLEAPAACTPPAPERPAELIVLSARTASALDAQAARLDHLETPSPQC 497
QY      485 LGDVAFLATTRSAMEHRLAVALATSRBGLRALDAAAGQOTSPGAIVRSIADSRGKLAFL 544
Db      498 LGDVAFLATTRSAMEHRLAVALATSRBGLRALDAAAGQOTSPGAIVRSIADSRGKLAFL 557
QY      545 FTGQGAQQLGMRGLYDWSAFREAFDLCVLLFMQELDRPLREVMWMAEPASVDAALLDOT 604
Db      558 FAGQGAQVPMGGRGLMEWMAFRFETFDRCVLLFMQELDRPLREVMWMAEPASVDAALLDOT 617
QY      605 AFTQPALFTFEYALAAALMRSGVEBELVAGHSIGELVACVAGVSLDEADAFVLAARGRL 664
Db      618 AFTQPALFTFEYALAAALMRSGVEBELVAGHSIGELVACVAGVSLDEADAFVLAARGRL 677
QY      665 MQALPAGAMVSIAPADVAANAAPHAAVSIANAAPDQVITAGACQPVHAIAMAANA 724
Db      678 MQALPAGAMVSIAPADVAANAAPHAAVSIANAAPDQVITAGACQPVHAIAMAANA 737
QY      725 RGARTKALVSHAFHSPLMAEPLAFGRVAVASVSRYRSTYLVNLSKCACTDEVSAGY 784
Db      738 RGARTKALVSHAFHSPLMAEPLAFGRVAVASVSRYRSTYLVNLSKCACTDEVSAGY 797
QY      785 WVRHAREVVRFADVQKALHAAAGAGTFVEVGKSTLLGLVAPACMPDAPALLAASRAGRDE 844

```

```

Db 798 WVRHAREVADGKALHAAAGLFVGVGPKFTLLGLVPACLPAPRVLLPASRAGDE 857
Oy 845 PATYLEALGMAVAGLVSMAGLFPSPGGRVLPFTYQORRYM-----DTKADDAAR-- 898
Db 858 AASALIEALGFWVVGGSVTMSGVFPSPGGRVLPFTYQORRYMIEAVDREAGDTGAR 917
Oy 899 ----- 898
Db 918 AGHPHLLGEVPSVTHAGLRIMETTLDRKRLPMIGEHRAQGVVPPGAGYLEMALSSGAE 977
Oy 899 -----GRRAP----- 904
Db 978 ILGDGPLOVTVLLETJTFAGDTAVPVQVYVTEERPGRLRFQVASREPGERRAPFRITHA 1037
Oy 905 ----- 904
Db 1038 RGVLRIGRVETPARSNLALPARLHAAVPAALYGALENGLOYGPAIRGLAELMRGEG 1097
Oy 905 ----- 904
Db 1098 EALGRVRLPEAAGSATAYQLHPVLLDACVMIVGAFADRDERTPWAVGVSVLFGQSP 1137
Oy 905 ----- 904
Db 1158 GELMCHARVSDGOASSRMSADELMDGTGAVVAEISRLVVERLASGVRRDADDMFLE 1217
Oy 905 -----GAGH----- 908
Db 1218 LDMEPALGPKITAGRWLLIGEGGGLRSLCSALKKAQHVVHAAAGDSTAGKRALLA 1277
Oy 909 ----- 908
Db 1278 NAFPGQAPTAVVHLSLDDGGQLGPGGLGAQGLADAPRPDVADALBESALRGCDSVLST 1337
Oy 909 ----- 908
Db 1338 VQALVGMDLRNAPRLMLTRGAQAAAAGVSVQAPLLGLRTTALBHAELRCISVDLP 1397
Oy 909 -----DEVEEGAVRGGDRRSARLDHPSPSGRREKVEAGDPRFLEI 952
Db 1398 AEPGEADALIAELLDAEEVVALRGSDRLVARLVHMLPAQOREKVEPGDPRFLEI 1457
Oy 953 DEPGVLDHLVLRVTERRAPGLGEVEIADVAGLSENDVQALAGWPDLPCKPMPPLLG 1012
Db 1458 DEPGALDQVLRATGRRAPGGEVEISVEAAGLDSIDQLALGVAPNDLPGEETIEPLVLS 1517
Oy 1013 GECAGRIYAVEGVNGLVGQPVIALSAGAFATHTTSAALVLPPOALSALIAAAMPVA 1072
Db 1518 SECAGRIYAVEGVNGLVGQPVIALAGVPATHTTSAATLVLPPLGLSATEAAMPVA 1577
Oy 1073 YLTMYMALDRIARLOPGRVLIHAATGVGLAAYOMAGHGAETHAATGTBEKAYLESL 1132
Db 1578 YLTMYMALDRIARLOPGRVLIHAAGGVGLCAVARMARVABEYATDTEENRAYLESL 1637
Oy 1133 GVRVYSDRSRDFVADVAMTGGEGVDVNLNLSGELIDKSFNLLRSHGRFVLEKRDY 1192
Db 1638 GVRVYSDRSRDFVADVAMTGGEGVDVNLNLSGELIDKSLMWLRACGLVKLGRDDC 1697
Oy 1193 ADNOLGRLPPLRNTSFSLLVDLRGMWLERPARVALLBEELGLIAGV----- 1239
Db 1698 ADTQPGLPPLRNTSFSQVDRGMWLDQPARIRALLDELFGVVAAGALSPLGSLRVGGS 1757
Oy 1240 FTTPPIATLPIYADADAFRMAOAGHLGKLVLTGLDPEVOQIRIPHA----- 1286
Db 1758 LTPPEVETFPISRAAEAFRRMAOGHGLKLVLTDDPEVRIRAPAESSVAVRADGTLYLT 1817
Oy 1287 -----GAG----- 1289
Db 1818 GGLGLGLRVAGWLAERGAQLVVGSRGAASAEGRAAVALLEAHGARVTYAKADVADRS 1877
Oy 1290 ----- 1289

```

```

Db 1878 QIERVLEVTASGNPLRGVVAAGLVDDGLMOQTPARFRTWGPVKVQGLHLHTLREA 1937
Oy 1290 ----- 1289
Db 1938 PLSPFVLYASAGLFGSPGQGNVAAAAPLDAISHHRAQGLPALSTDMKFTVGMAYA 1997
Oy 1290 -----PST----- 1292
Db 1998 QENRGARQISRGNGRITPDEGLSALARLLEGRVQGVIPITPQWVEFYATASRLS 2057
Oy 1293 -----GDDDLDRLASAPARAALAEFLRTQVSOVLRTEIKYGAELF 1338
Db 2058 RLVTQRAVDRAAGDDBLLEQLASAPASAGLLOVVRVQVSHVRLPEDKLEVAAPL 2117
Oy 1339 TRLGMDSLMAVELNRIEASIKLSTFTSTSPNIALAQNIL-DALATPL 1389
Db 2118 SSMGMDLSLELNRILEALGVAAPALGWTFTYVAALTRWLDLALVRL 2169

RESULT 13
US-09-568-472-7
; Sequence 7, Application US/09568472
; Patent No. 638719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568, 472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-7

Query Match 57.4%; Score 4137; DB 4; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8.5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

Oy 6 IERAEDPIAIVGASCRLLPGGVLDLSGFMTLLEGSRTDVGVPARMDAAMPDPDPAP 65
Db 26 LEQERTPIAIVGICGRFPGGADTRPAFWELLDGGRV--QPLDRMALGVHPSSEVP 83
Oy 66 GKTPVTASPLSD- VACPDASFEGISPREALRMDPAHRLLEVCWEALENNAIIPSAVLG 124
Db 84 -----RWAGLITAEVADGFDAFFGTSPREARSIDPQORLLLEVTEGLEADGIAPOSIDG 138
Oy 125 TENGVFICIGSEVEEALPQATASAEIDAAGGCTMPSVAGRTSYALGRCVPAVDTA 184
Db 139 SRTGVFLGACSSDSHTVAOQRE-EGDAVDITGNTLSVAAGRLSYTLGQPCLTVDTA 197
Oy 185 YSSSLVAVHLACOSLRSGECSTALAGVSLMSPSTLWLSKTRALARDCRKAFASEAD 244
Db 198 CSSSLVAIHLACRLRARRESLDALAGVNMLLSKTMIIGRIQALPDPGCRFPDASAN 257
Oy 245 GFGREGCAVVLKRLSGAPADGDRILAVIRGSAINHDAAGSGITVNGSSQIETVKRAL 304
Db 258 GFVREGCGMVVLKRLSDAQRHGDRIWALIRGSAMODGSGTGMADNVLAQEALEAL 317
Oy 305 ADAGCAASVYGVAHGTGTLGDPRIEIOALNAYGGRVAPLTLGSVYKTNLGHPEYA 364
Db 318 OSARVNDGALGYETHGTGTLGDPRIEVEALRAVLGPARADGSCVCAVYKTNLGHLEGA 377
Oy 365 SGITGLKVVLSIQHGQIPALHQAOLNPRIISWGLRLTVTRATPWPDMNTPRAGVSS 424

```



Beet Local Similarity 43.4%; Pred. No. 8.5e-297;  
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

```
QY 6 IERAADPIAVGASCRLLGVILDSGFMTLEGSRLDVGVPAREMDAAMPDDPAP 65
Db 26 LEOERTEBIAVIGICRPPGADTPPEARWEILDSGRDAV--QPLDRMALGVHSEVP 83
QY 66 GKTETVTRASFLLD-VACFDASFPGISPREALRMDPAHRLLEVCWEALENAIAPSLVG 124
Db 84 -----RNAGLITAEVDFDAFPGTSPREARSLDPQRLLELEVWEGLEDGIAFQSLDG 138
QY 125 TETGVFTIGIPSEYEALPQATASAEIDAHGGLGTMBSVGAGRISYALGRCPCVAVDTA 184
Db 139 SRTVFLGACSSDSYHTVAQQRRE-EQDAYDITGNTLSVAAGRLSYTLGLGCPCLTVDTA 197
QY 185 YSSSLVAVHLACOSLRGECSTALAGVSIMLSPSTJVLWLSKTLALADRGCKAFSAAD 244
Db 198 CSSLVAIHLACRRLARESDLALAGVNMLLSKTMTLGRIOALSPDGRCTFDASAN 257
QY 245 GFGREGCAVVVLRLLSGARADGRIILAVIRGSAIINHGAASSGLTVPGSSQEIYLRAL 304
Db 258 GFVAGGCGMVVLRLLSDQGHGRIMALLIRGSAMNOGRSTGLMARPVTLAQELALREAL 317
QY 305 ADAGCAASSVGYVEAHGTGTTLDPIETQALNAVYGLRDVATPLLIGSVKTLGHPRYA 364
Db 318 QSARVDAGALGCVETHGTGTSIGDPIEVEALRAVLGPARADGRCVLCGAVKTLGHEGA 377
QY 365 SGIRGLKLVLSLHGQIPALHQAOLNPRISWMDLRLTVRARTPMDMTTPRAGSS 424
Db 378 AGVAGLILKUALALHHELIPLHLPHHTLMPRIEQTALATEPVMPWRACRPPAGISA 437
QY 425 FGMGSGTAHVLEERAPATCTPPAPERPAELVLASRTASALDAQAARLDHLETYSQC 484
Db 438 FGLSGTIVHVLLEBPATVLAIPATPGSAELVLVSASAAALDQAARLSHIAHYEPQG 497
QY 485 LGDVAFSLATTRSAMEHRLAVALTSRGLRALDAAAQGTSPGAVSIADSSRGKLAFL 544
Db 498 LGDVAFSLVSTRSPMEHRLAVALTSREALALEVAAGQGTPGAARGRAASSPGKLAFL 557
QY 545 FTGGGAQTLCMGRGLYVWMSAFREAFDLCVTLFNOELDRPLREYVMAKPAVDAAALDQT 604
Db 558 FAGGAQVPGMGRGLMEAMPARFRETFCVTLFDRLEHQLPCEVMMAPGSSRSSLLDQT 617
QY 605 AFTOPALFTFEYALAAALWRSWGVPELVAGHSIGELVAACVAGFSLIEDAVFLVAARGRL 664
Db 618 AFTOPALFALEYALAAALFRSWGVPELVAGHSIGELVAACVAGFSLIEDAVFLVAARGRL 677
QY 665 MQALPAGGAMVSIAPADVAANAAPHAASVSIANAAPDOVTLAGAQCPVHAIAAAMA 724
Db 678 MQALPAGGAMVSIAPADVAANAAPHAALVSIANAAPVNGPEQVTLAGEKFPQOIAAFAA 737
QY 725 RGATATKLHSHAHSHPLMAPMLFAFGRVAVSVSYRRRSYLVNLSKACTDEVSSPGY 784
Db 738 RGATATKPLHSHAHSHPLMDPMLFAFRVTVSVYRRPSIALVNTLSKPCCTDEVSAPGY 797
QY 785 WVRHAREVVRPADGVKALHAAGAGTFVEVGKSTLLGLVPACMPDARPALAASSRAGDE 844
Db 798 WVRHAREVVRPADGVKALHAAGAGTFVEVGKPTLLGLVPCLEDPARVLLPARSAGDE 857
QY 845 PATVLEALGGLMAVGLVMAGLFPSSGGRVPLPTTYPMQORRYWI-----DTKADAAR-- 898
Db 858 AASALEALGFMVWGVGVTMGVFPSSGGRVPLPTTYPMQORRYWIEAVVREADGOTGAR 917
QY 899 ----- 898
Db 918 AGGHPLLGEVFSVSTHAGRLMETTLDRKLLPMLGHRAGGEVVPFGAGYLEMALSSGAE 977
QY 899 -----GDRRAP----- 904
Db 978 ILGDCPIQVTVDVLIETLTFAGDRAVVPQVVTTERPQRLAFQVARSREPGRRAPFRHA 1037
QY 905 ----- 904
```

```
Db 1038 RGVLRRIGRVETPARSNLALRARLHAAPAAIYGALENGIQYGPALRGALAEIMRGE 1097
QY 905 ----- 904
Db 1098 BALGRVLRPEAAGSATAYQLHPVLLDACVQMIYGAFAADREATPMAVEVGSVRLFORSP 1157
QY 905 ----- 904
Db 1158 GELWCHARVVSDDQOASSRWSADPELMDGTGAVVAEISRLVVERTLAGVRRRADDWPLE 1217
QY 905 -----GAGH----- 908
Db 1218 LDWEPAALGPKITAGRMILLGEGGLGRSLCSALKAAGHVVAHAADDSTAGMRALLA 1277
QY 909 ----- 908
Db 1278 NAFDQAPTAHVHLSSLDGGGQLGPGLAGAGALDAPSPDVVDALASALMRGCDVSL 1337
QY 909 ----- 908
Db 1338 VQALVGMDLRNAPRLMLTTRGAQAAAAGVSVVQAPLLGIGRTIALBHAELRCISVDLP 1397
QY 909 -----DEVEGCAVRCGDRRSARLDHPPESSGRREKYEAGDRPRLFI 952
Db 1398 AEPEGEADALLAEILLADDAEBEVALRGDRLVALVRLDPAQORREKVEPAGDRPRLFI 1457
QY 953 DEPGVLDHLVLRVTERAPGLGEVEIYVDAAGLSFNVQOLAGVPPDLPFGKMPRLTG 1012
Db 1458 DEPGALQVLVRATGRAPGGEVEISVEAAGLDSIDIQLALGAPNDLPBELEIPLVLG 1517
QY 1013 GECAGRIIVAGEGVNGLVGOPIVATASAGAFATHVTTSAALVLRPQALSAIEAAMPVA 1072
Db 1518 SEGAGRIIVAGEGVNGLVGOPIVATLAAGVATHVTTSATVLRPGLSATEAAMPVA 1577
QY 1073 YLTWYALDRIARLQPEBERVLIHAATGCVGLAAVQMAQHYGAETHARTGTEKAYLESL 1132
Db 1578 YLTWYALDRIARLQPEBERVLIHAAGGVGLCAVMAQORVGAETATADPENNAIYESTL 1637
QY 1133 GVRVYSDSRSDPFAVADVRAWTGCEGVNVVLSLSELIDXSFNLRSGHGRFVELGKDCY 1192
Db 1638 GVRVYSDSRSGRFTVDHANTGDEGVNVVLDLSGEBIDLSLWMLRACGRVLKLGRRDC 1697
QY 1193 ADNQLGLRPFLRNLSFSLVDLRCMMLERPARVRALLBEELGLLIAGV----- 1239
Db 1698 ADTQPGRLPRLRNLSFQOVLDRGMMLDQPARIRRLDELREGLVAAGAISPLGSLRVGGS 1757
QY 1240 FTTPPIATLPIARVADAFRSMAQOHLGKLVTLGDPEVOIRIPTHA----- 1286
Db 1758 LTPPEVETFPISRAAEFRMAQOHLGKLVTLGDPEVIRIRAPAESSVAVRADGYLV 1817
QY 1287 -----GAG----- 1289
Db 1818 GGLGGLGLRVAAGWLAEAGAQVLVGRSGAASAEORAAVAALEAHGARLVYAKADVADS 1877
QY 1290 ----- 1289
Db 1878 QIERVAREVTRASGMPLRAGVVAAGLVDDGLMOQTPARFRTVMGPXYOGALHTLTREA 1937
QY 1290 ----- 1289
Db 1938 PLSEFVLYASAAGLFGSPGOQNYAANAFLDALSHHRAOGLPALSIDMGFEVGNAVA 1997
QY 1290 -----PST----- 1292
Db 1998 QENRGARQISRMKRGITPPDEGLSALARLLEBGRVQTCVITPTPQWVEFPATASRRLS 2057
QY 1293 -----GDRDLRLASAAAPARAALAEFLRTOVSQVLRTEIKVGAELF 1338
Db 2058 RLVTTRQAVADRTRGDDDLBQLASAPSAAGLDQVAVVQVSHVRLRPEDKLEVADAPL 2117
QY 1339 TRIGMDSLMAVELRNRIEASIKLSTFTPLSTSNIALLAONLL-DALATRL 1389
Db 2118 SSMGMDLSMSELERNRIEALAGVAAAPALGWTYPTVAATIRMLDLDLVRL 2169
```

RESULT 15  
US-09-413-814-70  
Sequence 70, Application US/09413814  
Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Beyer, Stefan  
APPLICANT: Bioecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hoffe, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or  
FILE REFERENCE: PCT/US 99/2335  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 70  
LENGTH: 2259  
TYPE: PR1  
ORGANISM: Sorangium cellulosum  
US-09-413-814-70  
Query Match 53.1%; Score 3811.5; DB 3; Length 2259;  
Best Local Similarity 43.2%; Pred. No. 36-274;  
Matches 859; Conservative 118; Mismatches 245; Indels 767; Gaps 9;  
168 ISYALGRCPCVAVVTAYSSSLVAHVLACQSLRSGECSTALAGCVSLMSPETLWLSKT 227  
1 MSYTLGCGPLCTVIVICSSSLVAHILACRSRLRRESDTLALAGVMMLSSTMTMLGRI 60  
228 RLALADGRCKAFSADGRCGCAVVVLRKLSGARADDRILAVIRGSAINHGAASSG 287  
61 QALSDGHCRTDADANGVVRGCGGVVLRKLSGARADDRILAVIRGSAINHGAASSG 120  
288 LTVNGSSQEIYLRKALADAGCAASSVGVYEAHGTGTLGDIPIEQALNAVYGLGRDVA 347  
121 LMAPVLAQEAALRLALQARVDAGAGIGVETHGTGTLGDIPIEQALNAVYGLGRDVA 380  
348 PLLISVNTNIGHREYASITGLKLVLSLOHGOIPAHLLAQAALNPRISSWDLRTVTRA 407  
161 RCVLAVNTNIGHREYASITGLKLVLSLOHGOIPAHLLAQAALNPRISSWDLRTVTRA 440  
408 RTPMDWMTPRRAGVSSFGMSGTNHVVLEAPATCTPAPERPAELVLSARTASALD 467  
241 PVPWPRAGRRPRRAGVSSFGMSGTNHVVLEAPATCTPAPERPAELVLSARTASALD 500  
468 AQAARLRDHLLETYPQOCIGDVAFSLATTRSSAMEHRLVAATSRREGILRAALDAAGQSTP 527  
301 AQAARLRDHLLETYPQOCIGDVAFSLATTRSSAMEHRLVAATSRREGILRAALDAAGQSTP 560  
528 GAVRSIADSSRKLAFLLFGQAGQTLGMRGGLYDVMSAREAPDLCTVRLFNELDRPLRE 587  
361 GAARRAASSSPKLAFLFAGQAGQVFGMRGLMEAMPARETFDRCTVTLFDELHQPICE 620  
588 VMAAPASVADALLOTAFTOPALFTFEYALAAWRMSGVBEPLVAGHSIGELVACVAG 647  
421 VMAAPASVADALLOTAFTOPALFTFEYALAAWRMSGVBEPLVAGHSIGELVACVAG 680  
648 VESLEDAVFLVARGRLMQLPAGGAMVSIAPADVAAPAAVAPHAASVIAAVNAPDOVV 707  
481 VESLEDAVFLVARGRLMQLPAGGAMVSIAPADVAAPAAVAPHAASVIAAVNAPDOVV 740

708 IAGAGPVHAIATAAAGARTKALVSHASHPLMAPMLAEFGVAVESVRRPSTVAV 767  
541 IAGAGPVHAIATAAAGARTKALVSHASHPLMAPMLAEFGVAVESVRRPSTVAV 800  
768 SNLGGACTDVSPPGVVVRHAREVRPADGVKALHAAGAGTFEVEGPKSTLLGLVPAAM 827  
601 SNLGGACTDVSPPGVVVRHAREVRPADGVKALHAAGAGTFEVEGPKSTLLGLVPAAM 860  
828 PDARPALLASSRAGDEPATVLEALGIMAVGLVSNAGLTPSGRRVPLPTYPWQERY 887  
661 PDARPALLASSRAGDEPATVLEALGIMAVGLVSNAGLTPSGRRVPLPTYPWQERY 920  
888 WI-----DKADDAAR----- 898  
721 WIEAPVDEADGTGARRAGCHPLGCEVSVSTHAGRLMETTLDRKRLPMLGHRAGCEV 880  
899 ----- 898  
781 VEPGAGVLEMLSSGAEITLGDPIQVTDVVLITLTFAGDTAVPVQVVTTEBRPGLRFQ 840  
899 -----GRRAP----- 904  
841 VASREPGERRAPFRIHARGVLRIGRVETPARSNLALRARIHAAPAAIYGALAEML 900  
905 ----- 904  
901 QYPALRGLAELMRGEALGRVLRLEPAGSATAYQLHPVLLDACVQMTVGFADREAT 960  
905 ----- 904  
961 PMAPVGVSVRLPQSPPELWCHARVVSDDQASRMSADPELMDGTGAVVAEISRLVVE 1020  
905 -----GAGH----- 908  
1021 RLASGVRRRDDWFLBEMEPALGPKITAGRWLLLEGSGGLRSLCSALKAAHGVV 1080  
909 ----- 908  
1081 HAAGDSTAGMRALLANAPGQAFVAVVHLSLDGGQLGPGAGCALDAPNSPDVA 1140  
909 ----- 908  
1141 DALBSALMRGDSVLSVQLVGMDLRNPRLMLTRGAQAAAAGDVSVQAPLLGLRT 1200  
909 -----DEVEGCAVGRGDRSARLDHPPESSG 935  
1201 IALBEALRCISVDLPAPPEGEADALIAELLADDAEELVLRGDDRLVRLVHRLPDAQ 1260  
936 RREKVEAGDRPPRLTIDEPGLVLRVTERRAPGLGEVEIVDAAGLSFNDVQLALG 995  
1261 RREKVEAGDRPPRLTIDEPGLVLRVTERRAPGLGEVEIVDAAGLSFNDVQLALG 1320  
996 WVPDDLPGKPNPPLLIGECAGRIYAVGEGVGLVGOVITALSAGAFTHVTTSAALVL 1055  
1321 VAPNDLPPEBIEPLVIGSECAGRIVAVGEGVGLVGOVITALSAGAFTHVTTSAALVL 1380  
1056 PRPOLSLATEAAMPVAVLTMYALDRABKOPGERVLIHAATGVGGLAAVQMAOHVAG 1115  
1381 PRPOLSLATEAAMPVAVLTMYALDRABKOPGERVLIHAATGVGGLAAVQMAOHVAG 1440  
1116 VHAATGPEKAPVLESIGVRYVSDRSRDPADVAMTGGEGVDVVLNLSGELLDKSFN 1175  
1441 VHAATGPEKAPVLESIGVRYVSDRSRDPADVAMTGGEGVDVVLNLSGELLDKSFN 1500  
1176 LIRSHREVELEKRDVADNOLGLRPFRLNLSFSLVLDLGMMLERPARVRLLEBLGLI 1235  
1501 VLRACGRVYKGRDDCADTQGLRPLLRNLSFSGVVDLRGMWLDQPARIRLLDLFLGLV 1560  
1236 AAGV-----FTPPPIATLPYARVADAFRSMAQOHLCKVLTLTGDEPVQIRI 1282  
1561 AAGV-----FTPPPIATLPYARVADAFRSMAQOHLCKVLTLTGDEPVQIRI 1620  
1283 PTHA-----GAG----- 1289

```

Db 1621 PAESSVAVRADGYLVVTGGLGLGRVAGWLAERGAGQLVWGRGAASAEORAAVALE 1680
QY 1290 ----- 1289
Db 1681 AHGARVTVAKADVADRSQIERVLRBEVTASGMPLRGVVHAAGLVDDGLMOQTPARFTVM 1740
QY 1290 ----- 1289
Db 1741 GPKVQALHHTLTREAPLSFVLYASAAGLFGSPGQGNVAANAFLDALSHHRAOGLP 1800
QY 1290 ----- 1289
Db 1801 ALSIDWGFTEVGMAVAQENRGARQISKMRGITPDEGLSALARLLEGDRVQGVIPITP 1860
QY 1290 ----- 1321
Db 1861 RQWVEFPATIASRRLSLVTTORAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQV 1920
QY 1322 SOVLRTPEIKVGAELFTRLGMDSIMAVELNRRLEASIKLSTFFLSTSPNIALLAQNL 1381
Db 1921 SHVLRLEPEDKIEVDAPLSSMGMDSIMSELENNRIEALGVAPAPALGWTYFTVAALITRWL 1980
QY 1382 L-DALATL 1389
Db 1981 LDDALVVRL 1989

```

Search completed: October 2, 2003, 17:35:52  
 Job time : 71.7527 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 95.763 Seconds  
(without alignments)  
2355.298 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210  
1 VADPPIRAEDPIAIVGAS.....GVQNDPVSSGADQWETIAL 1421

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /Geneseq\_19Jun03:.\*  
2: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.\*  
3: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.\*  
4: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.\*  
5: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.\*  
6: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.\*  
7: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.\*  
8: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.\*  
9: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.\*  
10: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.\*  
11: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.\*  
12: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.\*  
13: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.\*  
14: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.\*  
15: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.\*  
16: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.\*  
17: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.\*  
18: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.\*  
19: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.\*  
20: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.\*  
21: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.\*  
22: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.\*  
23: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.\*  
24: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2002.DAT.\*  
25: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7210	100.0	1421	21	AAV58573	Sorangium cellulos
2	4137	57.4	2439	21	AAV58578	Sorangium cellulos
3	3311.5	45.9	7257	21	AAV58576	Sorangium cellulos
4	2531.5	35.1	2153	22	AA83974	Amino acid sequenc
5	2363	32.8	4038	22	ABG9871	S. cinamonensis M
6	2355.5	32.7	5435	22	AAE10145	Streptomyces nous
7	2351	32.6	4106	22	ABG9872	S. cinamonensis M
8	2331.5	32.3	11096	22	AAE10129	Streptomyces nous
9	2319.5	32.2	2149	24	ABP57679	Saccharopolyspora

10	2273	31.5	2152	20	AAV9298	SpnB a polyketide
11	2273	31.5	2152	22	AA870966	S. spinosa protein
12	2260.5	31.4	7068	22	AAE10142	Streptomyces nous
13	2260.5	31.4	9477	22	AAE10144	Streptomyces nous
14	2250	31.2	6095	22	AAU10701	Amino acid sequenc
15	2245.5	31.1	3798	21	AAV58577	Sorangium cellulos
16	2194	30.4	5087	22	AAU10700	Amino acid sequenc
17	2165	30.0	6797	22	AA831558	Pimaricin biosynth
18	2150	29.8	1841	18	AAW22605	Tylosin synthase
19	2135	29.6	4132	22	ABG9870	S. cinamonensis M
20	2128	29.5	3816	21	AAV92708	S. antiocticus 8'
21	2127	29.5	3170	20	AAV9299	SpnC a polyketide
22	2127	29.5	3170	22	AA870967	S. spinosa protein
23	2125	29.5	4630	21	AAV77177	S. venezuelae vep
24	2119	29.4	4933	24	ABP57681	Saccharopolyspora
25	2113	29.3	4924	22	AA870968	S. spinosa protein
26	2113	29.3	4928	20	AAV93300	SpnD a polyketide
27	2108.5	29.2	3562	22	AA882213	Polyketide synthas
28	2107.5	29.2	3567	14	AA844431	eryA region polype
29	2104	29.2	2238	22	ABG9869	S. cinamonensis M
30	2103	29.2	4630	18	AAW19629	Streptomyces venez
31	2097.5	29.1	1366	22	AAE10137	Streptomyces nous
32	2097	29.1	3167	24	ABP57680	Saccharopolyspora
33	2091.5	29.0	5588	20	AAV93301	SpnE a polyketide
34	2091.5	29.0	5588	22	AA870969	S. spinosa protein
35	2078	28.8	4344	24	ABP57678	Saccharopolyspora
36	2065.5	28.6	5564	24	ABP57682	Saccharopolyspora
37	2063.5	28.6	3413	19	AAW52849	A. mediterranei ri
38	2052.5	28.5	3739	21	AAV77193	S. venezuelae macr
39	2052.5	28.5	3739	21	AAV77201	S. venezuelae pik
40	2052.5	28.5	12199	21	AAV77180	S. venezuelae pik
41	2051.5	28.5	4551	21	AA818637	Amino acid sequenc
42	2051.5	28.5	4551	21	AAV67201	Narboxilide syntha
43	2051.5	28.5	4551	24	ABG71661	S. venezuelae narb
44	2051.5	28.5	4613	21	AAV77192	S. venezuelae macr
45	2051.5	28.5	4613	21	AAV77200	S. venezuelae pik

#### ALIGNMENTS

RESULT 1	AAV58573	standard; Protein; 1421 AA.
ID	AAV58573;	
XX	AC	
XX	AAV58573;	
DT	10-APR-2000	(first entry)
DE	Sorangium cellulosum type I polyketide synthase EPS A.	
KW	EPDS A, type I polyketide synthase; epothione biosynthesis;	
KW	thiazole ring formation; taxol substructure; anticancer.	
OS	Sorangium cellulosum,	
XX		
XX	MO9966028-A2.	
PD	23-DEC-1999.	
XX		
PF	16-JUN-1999;	99WO-EP04171.
XX		
PR	18-JUN-1998;	98US-0099504.
PR	24-SEP-1998;	98US-0101631.
PR	05-FEB-1999;	99US-0118906.
XX		
PA	(NOVS ) NOVARTIS AG.	
PA	(NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.	
PI	Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;	
DR	WPI; 2000-097741/08.	
DR	N-PSDB; AA255887.	

XX New isolated epothilone synthase genes, used for the recombinant  
 PT production of epothilone for use in cancer therapy

Claim 12; Page 105-109; 174pp; English.

CC This sequence represents a Sorangium cellulosum type I polyketide  
 CC synthase, EPOS A, which is one of several epothilone biosynthetic  
 CC enzymes encoded by a 68.75 kb contig. Epothilones A and B are  
 CC 16-membered macrocyclic polyketides with an acylcysteine-derived  
 CC starter unit; polyketides being synthesised from two-carbon building  
 CC blocks, the beta-carbon of which always carries a keto group. Each round  
 CC of two-carbon addition is carried out by a complex of enzymes known as  
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.  
 CC EPOS A (AA158573) and EPOS P (AA158574) are involved in formation of  
 CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D  
 CC and EPOS E (AA158575-158578) are involved in polyketide backbone  
 CC formation. EPO F (AA158579) is an epothilone macrolactone oxidase, and  
 CC the proteins Orf 3 (AA158582) and Orf14 (AA158593) are thought to be  
 CC involved in transport. Epothilones mimic the biological activity of  
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic  
 CC compositions. Epothilones exhibit a much lower drop in potency against a  
 CC multiply drug-resistant cell line compared with taxol, and are  
 CC considerably less efficiently exported from such cells by the multidrug  
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of  
 CC epothilones as anticancer agents, they are problematical to produce on a  
 CC large scale. Epothilones are too complex for industrial scale chemical  
 CC synthesis, and Sorangium cellulosum is difficult to ferment, producing  
 CC poor yields of epothilones. The nucleic acids of the invention may be  
 CC used for the recombinant production of epothilones in a heterologous host  
 CC that is more amenable to fermentation.

Sequence 1421 AA;

Query Match 100.0%; Score 7210; DB 21; Length 1421;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRIPIERAADPIAIVGASCRIPGVIDLISGFWTLLEGSDRTVGVRPERMDAAWFP 60  
 DB 1 VADRIPIERAADPIAIVGASCRIPGVIDLISGFWTLLEGSDRTVGVRPERMDAAWFP 60  
 QY 61 DDDAGKPTVTRASFSDVACFDASFPFISPREALRMDPAHRLLEVCHEALLENALIAS 120  
 DB 61 DDDAGKPTVTRASFSDVACFDASFPFISPREALRMDPAHRLLEVCHEALLENALIAS 120  
 QY 121 ALVGETGVFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGLRCPVA 180  
 DB 121 ALVGETGVFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGLRCPVA 180  
 QY 181 VDTAVSSSLVAHVLACQSLRSGECSTALAGVSLMSPESTLYWLSTKTRALADGRCKAF 240  
 DB 181 VDTAVSSSLVAHVLACQSLRSGECSTALAGVSLMSPESTLYWLSTKTRALADGRCKAF 240  
 QY 241 AADAGRGREGGCAVYVLRKLSGARADGRIILAVIRGSAINHPGASGLVPPGSSQEIYL 300  
 DB 241 AADAGRGREGGCAVYVLRKLSGARADGRIILAVIRGSAINHPGASGLVPPGSSQEIYL 300  
 QY 301 KBALADAGCAASSVGVVEAHGTGTLGDIIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360  
 DB 301 KBALADAGCAASSVGVVEAHGTGTLGDIIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360  
 QY 361 PEYASGITGLKRVYVSLQHGQIPAHILHAQALNPRISWGLRLTVTRARTPMPDWTTPRA 420  
 DB 361 PEYASGITGLKRVYVSLQHGQIPAHILHAQALNPRISWGLRLTVTRARTPMPDWTTPRA 420  
 QY 421 GVSSFGMSGTNAHVVEEAPATCTPPAPERPAEILVLSARTASALDAADARLRLHLEY 480  
 DB 421 GVSSFGMSGTNAHVVEEAPATCTPPAPERPAEILVLSARTASALDAADARLRLHLEY 480  
 QY 481 PSQCLGDVAFSLATTSAMEHRLAVALTSREGLRALDAAAGQTSPPGAVRSIADSSRCK 540  
 DB 481 PSQCLGDVAFSLATTSAMEHRLAVALTSREGLRALDAAAGQTSPPGAVRSIADSSRCK 540

QY 541 IAFLEFGGAGTGLGNGRLYDWSAFREAFDLCVLFENQELDRPULREVMMAEPASVDAAL 600  
 DB 541 IAFLEFGGAGTGLGNGRLYDWSAFREAFDLCVLFENQELDRPULREVMMAEPASVDAAL 600  
 QY 601 IDQTAFTQPALFTFEYALAAALMRSGVEPELVAGHSIGELVAAQVAFSLEDVAFVAA 660  
 DB 601 IDQTAFTQPALFTFEYALAAALMRSGVEPELVAGHSIGELVAAQVAFSLEDVAFVAA 660  
 QY 661 RGRIMQALPAGAMVSIAPADVAAVAAPHAASVIAAVNAPDOVYTAGQPVHAITA 720  
 DB 661 RGRIMQALPAGAMVSIAPADVAAVAAPHAASVIAAVNAPDOVYTAGQPVHAITA 720  
 QY 721 AMAAGATKTLHVAHSHSPMLMAEAFGRVNESVYRPSIYVNLGSKACTDEVS 780  
 DB 721 AMAAGATKTLHVAHSHSPMLMAEAFGRVNESVYRPSIYVNLGSKACTDEVS 780  
 QY 781 SPGYVWRARAREVRPADGVKALHAAGACTFVEVGPKSTLGLVPACMPDAPALIASRA 840  
 DB 781 SPGYVWRARAREVRPADGVKALHAAGACTFVEVGPKSTLGLVPACMPDAPALIASRA 840  
 QY 841 GRDEPATVLEALGLMAVGLVSWAGLFPSCGRVPLPTYWQREYWIITKADDAAGD 900  
 DB 841 GRDEPATVLEALGLMAVGLVSWAGLFPSCGRVPLPTYWQREYWIITKADDAAGD 900  
 QY 901 RRAFGAGHDEVVEGAGVAGGDRSRRLDHPPESSRREKVEAAGRPRLTIDEGVLDH 960  
 DB 901 RRAFGAGHDEVVEGAGVAGGDRSRRLDHPPESSRREKVEAAGRPRLTIDEGVLDH 960  
 QY 961 LVLRTERARRPCLGSEVEIAYDAAGLSFNDVOLALGMVDDLPKGNPPLLGGECAGRIY 1020  
 DB 961 LVLRTERARRPCLGSEVEIAYDAAGLSFNDVOLALGMVDDLPKGNPPLLGGECAGRIY 1020  
 QY 1021 AVGEGVNLVWGQPVIALSAGAFATHTVTTSAALVLRPQALSAIEAAMPVAYLTAWYAL 1080  
 DB 1021 AVGEGVNLVWGQPVIALSAGAFATHTVTTSAALVLRPQALSAIEAAMPVAYLTAWYAL 1080  
 QY 1081 DRIARLOGERVLTHAAGVGVLAOVAOHVGAHVHATAGPEKRAVLESIGVYVSDS 1140  
 DB 1081 DRIARLOGERVLTHAAGVGVLAOVAOHVGAHVHATAGPEKRAVLESIGVYVSDS 1140  
 QY 1141 RSDRFVAIVRAMTGGEGVDVVLNLSGELIDKSFNILRSRHFVYELGRDVCYADNQLGR 1200  
 DB 1141 RSDRFVAIVRAMTGGEGVDVVLNLSGELIDKSFNILRSRHFVYELGRDVCYADNQLGR 1200  
 QY 1201 PELRNLSFSLVDLRGMLEPRARVALLIEELIGLIAAGVTPPIATLPIARVADAFPSM 1260  
 DB 1201 PELRNLSFSLVDLRGMLEPRARVALLIEELIGLIAAGVTPPIATLPIARVADAFPSM 1260  
 QY 1261 AQAQHLGKLVLTLDGPEVOIRPTHAAGAPSTGDDLDRLASAPAPARAAALEFRLTQ 1320  
 DB 1261 AQAQHLGKLVLTLDGPEVOIRPTHAAGAPSTGDDLDRLASAPAPARAAALEFRLTQ 1320  
 QY 1321 VSQVARTPEIKVGALELFTRLGMSLMAVELRNRIEASLKLKLTSTFTSTPNIALAQN 1380  
 DB 1321 VSQVARTPEIKVGALELFTRLGMSLMAVELRNRIEASLKLKLTSTFTSTPNIALAQN 1380  
 QY 1381 LLDALATLSLERRVAENLRAGVONDFVSSGADQDWEIITAL 1421  
 DB 1381 LLDALATLSLERRVAENLRAGVONDFVSSGADQDWEIITAL 1421

RESULT 2

AA158578  
 ID AA158578 standard; Protein; 2439 AA.

AA158578;

DT 10-APR-2000 (first entry)

DE Sorangium cellulosum type I polyketide synthase EPOS E.

KW EPOS E; type I polyketide synthase; epothilone biosynthesis;

KW polyketide backbone formation; taxol substitute; anticancer.  
 XX Sorangium cellulosum.  
 XX MO9966028-A2.  
 XX  
 PD 23-DEC-1999.  
 XX  
 PF 16-JUN-1999; 99MO-BP04171.  
 XX  
 PR 18-JUN-1998; 98US-0099504.  
 PR 24-SEP-1998; 98US-0101631.  
 PR 05-FEB-1999; 99US-0118906.  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 PI Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;  
 DR WPI; 2000-097741/08.  
 DR N-PSDB; AA255887.  
 XX  
 PS New isolated epochohione synthase genes, used for the recombinant  
 PT production of epochohione for use in cancer therapy .  
 XX  
 XX Claim 12; Page 151-158; 174pp; English.  
 CC This sequence represents a Sorangium cellulosum type I polyketide  
 CC synthase, EPOS E, which is one of several epochohione biosynthetic  
 CC enzymes encoded by a 68.75 kb config. Epochohiones A and B are  
 CC 16-membered macrocyclic polyketides with an acylcysteine-derived  
 CC starter unit; polyketides being synthesised from two-carbon building  
 CC blocks, the beta-carbon of which always carries a keto group. Each round  
 CC of two-carbon addition is carried out by a complex of enzymes known as  
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.  
 CC EPOS A (AAVS58573) and EPOS P (AAVS58574) are involved in formation of  
 CC the thiazole ring formation of epochohiones, and EPOS B, EPOS C, EPOS D  
 CC and EPOS E (AAVS58575-Y58578) are involved in polyketide backbone  
 CC formation. EPO F (AAVS58579) is an epochohione macrolactone oxidase, and  
 CC the proteins Orf 3 (AAVS58582) and Orf14 (AAVS58593) are thought to be  
 CC involved in transport. Epochohiones mimic the biological activity of  
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic  
 CC compositions. Epochohiones exhibit a much lower drop in potency against a  
 CC multiply drug-resistant cell line compared with taxol, and are  
 CC considerably less efficiently exported from such cells by the multidrug  
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of  
 CC epochohiones as anticancer agents, they are problematical to produce on a  
 CC large scale. Epochohiones are too complex for industrial scale chemical  
 CC synthesis, and Sorangium cellulosum is difficult to ferment, producing  
 CC poor yields of epochohiones. The nucleic acids of the invention may be  
 CC used for the recombinant production of epochohiones in a heterologous host  
 CC that is more amenable to fermentation.  
 XX  
 SQ Sequence 2439 AA;  
 Query Match 57.4%; Score 4137; DB 21; Length 2439;  
 Best Local Similarity 43.4%; Pred. No. 0;  
 Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

Db 198 CSSSLVAIHLACRSRLRRESLDALAGVNMILSSKTMIMLGRIGALSBDGCRTPDASAN 257  
 QY 245 GFGREGCAVVLKRLSGARADGDRILAVIRGSAINHDSAGSLTPNGSQEIVLKRAL 304  
 Db 258 GFVREGCGAVVLKRLSDAQHGRIRWALIRGSAAMNDGRTGLMAPRVLAQELRLRAL 317  
 QY 305 ADAGCAASVGVYEAHGTTTIGDPIETIOALNAVYGLRDVATPLLIGSVKTLGHPEYA 364  
 Db 318 QSARVDAGALIGVETHGTGTSLGDPTEVEALRAVLGPBARADGSRCLCAVXNLIHLEGA 377  
 QY 365 SGITGLKLVLSLHQGQIPALHQAOLNPRISMGDKLTITVTRATTPWDMNTPPRAGISS 424  
 Db 378 AGVAGLTKAALALHRELIPRLHFTINPRLIRIGETALALATEBPVMPRAGRPRAGISA 437  
 QY 425 FGMSGTAHVVLLEAPATCTPPAPERPAELVLSARTASALDQAARLRHOLETPSQ 484  
 Db 438 FGLSGTNHVHVLLEAPATVLAATPGSABELLVLSASAAALDQAARLSHIAIYPPQG 497  
 QY 485 LGDVAFSIATTRSAMEHRLAVALAATSRGLRAALDAAAQGTSPGAVRSIADSRGKLAPL 544  
 Db 498 LGDVAFSIATSTRSMEHRLAVALAATSRGLRAALDAAAQGTSPGAAARGAASRSGKLAPL 557  
 QY 545 FTGGAGTLLGMRGLYDVMSAFREAPDLCVLFPQELDRPLREVMMAEPASVDAALLDQT 604  
 Db 558 FAGGAQVPPGKGLMEAMPARFETPDRCVTLFPRELHQPCEVWMAEPGSSRSSLLDQT 617  
 QY 605 AFTQPALTFEYALAAALMRSGVPELVAAGHSIGELVACVAGVFSLEDAVFLVAARGRL 664  
 Db 618 AFTQPALFALEYALAAALFRSGVPELVAAGHSIGELVACVAGVFSLEDAVFLVAARGRL 677  
 QY 665 MQALPAGGAWVSIAPADVAVAAPPAASVSIANAAPDQVITAGACQPVHAIAMAAMA 724  
 Db 678 MQALPAGGAWVSIAPADVAVAAPPAALVSIANVGPQVITAGAEKFPQVIAAIPAA 737  
 QY 725 RGATTKALHVAHSHPSPLMAEMLEAFGRVAESVYRRPSIVLVNLSKACTDEVSSTGY 784  
 Db 738 RGATTKPLHVAHSHPSPLMDMLAFRVTYSVYRRPSIALVNLSSKPTCTDEVSATGY 797  
 QY 785 WVRAREVVRADQVKALHAAGAPFVEGKSTLLGLVPACMPDPARALAAASRAGDE 844  
 Db 798 WVRAREVVRADQVKALHAAGAPFVEGKSTLLGLVPACMPDPARALAAASRAGDE 857  
 QY 845 PATVLEALGSLMAVAGVLSMAGLPPSGGRVPLPTYPQWRBYI---DTKADAAR-- 898  
 Db 858 AASLALBAGVWVGVSTAGVPPSGGRVPLPTYPQWRBYI---DTKADAAR-- 917  
 QY 899 -----GRRAP----- 904  
 Db 978 ILGDPPIQVTDVLIETLTPAGDTAVPVQVTTTEERPGRLFOVASRPEGRRAPFRITHA 1037  
 QY 905 ----- 904  
 Db 1038 RGVLRIGRVETPARSNLALRARLHAAPPAATYGALEMGLOYPALRGALBELMRGEG 1097  
 QY 905 ----- 904  
 Db 1098 EALGRVLRPEAASATAYQALPVLLDACVOMIVGAFADRDDEATPMAPVGVSVLFRSP 1157  
 QY 905 ----- 904  
 Db 1158 GELMCHARVVSDDQOASRMSADPELMDGTAVVAEISRLVVERLASGVRRRDADWLE 1217  
 QY 905 -----GAGH----- 908  
 Db 1218 LDWEPPALGSPKITAGRWLLLGEGGGLGRSLCSALKAAGHVVAAGDDTSTAGWRALLA 1277  
 QY 909 ----- 908  
 Db 1278 NAFDGAQPTAVVHLSSLDGGQLPGGLAOGALDAPRSFVDVADALBSALMRGCDSVLSL 1337

```

QY 909 ----- 908
Db 1338 VOALVGMRLRNAPRLMLLTRGAQAAAAGDVSVQAPLLGLGRTIALENHAEIICISVDLP 1397
QY 909 ----- DEVEEGGAVRGRRSARLDHPPESSGRRREKVEAAGDRPRLRI 952
Db 1398 AEPGEADLALAEADDAEEVALLRGGRRLVRLVHRLPDQORREKVEAAGDRPRLRI 1457
QY 953 DEPGVLDHLVLRVTERBARGLGEVEIADVAAGLSFNDVOLALGMVDDLPGRKPNPPLL 1012
Db 1458 DEPGALDQVLRLATGRRAPGGEVEISVEAAGLSIDIQALGVANPDLPGEEIIEPLV 1517
QY 1013 GEGCARIVAVGSGVGLVVGQVIALSAGAPATHYTSALVLRPQALSAIEAAMPRA 1072
Db 1518 SECARIVAVGSGVGLVVGQVIALSAGAPATHYTSALVLRPQALSAIEAAMPRA 1577
QY 1073 YLTAWYALDRIRARLOPGEVRLIHAATGVLGAQVQAQVAVHATAGTPEKRAVLESL 1132
Db 1578 YLTAWYALDKVAHLQAGEVRLIHAAGVGLCAVRAQVGAEVATADTPENRAVLESL 1637
QY 1133 GVRVYVSDSSDRFVADVRAWTGEGVDVVLNSLSELIDKSFNLRSHGRFVELGRDXY 1192
Db 1638 GVRVYVSDSSGRFVTDVHAMTDEGVVDVLDLSGERIDKSIIMVLRAQGLVLTGRDDC 1697
QY 1193 ADNOLGRFRLNLSFLVLAGMMLERPARVRLLEILLGLIAGV----- 1239
Db 1698 ADTQPLPLRLNFSQVDLRGMMLDQPARIRALLDELFGVAAAGALSPLSGSLRGGS 1757
QY 1240 FTPEPIATLPARVADAFRMAQAQHLGKLVITIGDPEVQIRIPTHA----- 1286
Db 1758 LTRPEVETPISRAAEAFRRMAQGGHLGKLVITLDPEVRIRAPAESSVAVADGTIVLT 1817
QY 1287 -----GAG----- 1289
Db 1818 GGLAGGLGRVAGMLAERGAQGVLVGRSGAASAEQRAVAALAHGARVYAKADVAERS 1877
QY 1290 ----- 1289
Db 1878 QIERVIREVTASGMPLRGVVAAGLVDDGLMQCPARFRTVMGPKVQALHLHTLTREA 1937
QY 1290 ----- 1289
Db 1938 PLSFVLYASAAGLFGSGQGVYAAANAFLLDSHHRAGQGPALSIDKMTVEGMAYA 1997
QY 1290 ----- PST----- 1292
Db 1998 QENRGARQISRGMRGITPDEGLSALARLEEGDRVQGVIPITPQWVEFYPATASRRLS 2057
QY 1293 -----GDRDLDRLASAPARAALAEFLRTQVSQVLRPEIKVGAELF 1338
Db 2058 RLVTTQRAVADRTAGDRDLLEQALSAEPARAGLLQDVVRVSHVLRLEPEKIEVDAPL 2117
QY 1339 TRLAGDSLMAVELRNRIEASLKLSTPLSTSPNIALIAQNL-DALATAL 1389
Db 2118 SSMGMDSLMSLELRRIEALGVAAAPALGMWYPTVAATRLMLDADLVRL 2169

```

```

PN MO9966028-A2.
XX 23-DEC-1999.
PD 16-JUN-1999; 99WO-EP04171.
XX 18-JUN-1998; 98US-0099504.
PR 24-SEP-1998; 98US-0101631.
PR 05-FEB-1999; 99US-0118906.
XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Schupp T, Liqon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
DR WPI; 2000-097741/08.
DR N-PSDB; AA255887.
PT New isolated epothonione synthase genes, used for the recombinant
PT production of epothonione for use in cancer therapy
XX
PS Claim 12; Page 118-139; 174pp; English.
XX
CC This sequence represents a Sorangium cellulosum type I polyketide
CC synthase, EPOS C, which is one of several epothonione biosynthetic
CC enzymes encoded by a 68.75 kb contig. Epothoniones A and B are
CC 16-membered macrocyclic polyketides with an acylcysteine-derived
CC starter unit; polyketides being synthesized from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AA558573) and EPOS P (AA558574) are involved in formation of
CC the thiazole ring formation of epothoniones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AA558575-58578) are involved in polyketide backbone
CC formation. EPO F (AA558579) is an epothonione macrolactone oxidase, and
CC the proteins Orf 3 (AA558582) and Orf14 (AA558593) are thought to be
CC involved in transport. Epothoniones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epothoniones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epothoniones as anticancer agents, they are problematical to produce on a
CC large scale. Epothoniones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epothoniones. The nucleic acids of the invention may be
CC used for the recombinant production of epothoniones in a heterologous host
CC that is more amenable to fermentation.
XX
SQ Sequence 7257 AA;

```

Query Match 45.9%; Score 311.5; DB 21; Length 7257;  
Best Local Similarity 41.1%; Pred. No. 2.3e-254;  
Matches 784; Conservative 162; Mismatches 432; Indels 531; Gaps 25;

```

QY 9 AABDPITAVGASCRPLGGVVDLDSGFMTLLESRDVTGVAPARMDAAMPDPDPAPAKT 68
Db 3021 ASDEPIALVGAACRPPGVEDLESYTWQIAGGVVSAVAPDRMDAAWYDDPEITPRT 3080
QY 69 PVTASPLSDVACPDASFFGISPREALMDPAHRLLEVCWEALENAAIASALVGTETG 128
Db 3081 YVTKAFILRDLQRIDATFFRISPREAMSLDQOGRLLLEVSWALASAGIADPTLDSPTG 3140
QY 129 VFTIGSEVEYALPQATASAEIDAAGLGMTPSYGAARISYALGRPCAVADVATYSS 188
Db 3141 VFVGAPMEYTYQRLRGFTDGAAGLYGGTGMMLSVTAGRLSFFGLHGTPLAMDYACSS 3200
QY 189 LVAVHLACOSLRSGGCTALAGVSLMSPSTIWLSTRALARDGRKAFSAEADGFR 248
Db 3201 LVALLHACOSLRGLGCDALGVGVVLAPEFVLLSMRLASPDGRKTTSSADADGTAR 3260
QY 249 GEGCAVVVLKRLSGARADGRIILAVIRGSAINHDSAGLTVPNGSQEIVLKRALADAG 308

```

```

Db 3261 GEGCAVVVVKRLRDAQRAGDSILALIRGSANVHDESSGLTVNGPAQOALLRQALSOAG 3320
Qy 309 CAASSVGVYEAHGTGTTLDGPRIEIOALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGIT 368
Db 3321 VSPVDVDEVECHGCTALGDPIEYQALSEVYGPBGSGGRPLVIGAKAKNVNHLAASGLA 3380
Qy 369 GLIKVYLSLQGGQIPAHLLHAQALNFRISWGDLRLTVTRARTPMDMTPRRAGVSSFGMS 428
Db 3381 SLTAAVVALRHHQIIPAOQELGELNPHLPMNTLPVAVPRKAVPWGAGPRRAGVSAFGLS 3440
Qy 429 GTNAHVYVEEAPATCTEPAPERPAELIYLSARTSALDAAARLDLETFYPSQCIGDV 488
Db 3441 GTNAHVYVEEAPERPAERPAERPAERPAERPAERPAERPAERPAERPAERPAERPAERPA 3500
Qy 489 AFSIATTSAMEHRLAVATSRREGIRALDAAAOQOTSGAVRSIADSSRKLAFLEFTGQ 548
Db 3501 AFSIATTSRPMHRLAITSREALRGALDAAAOQKTQGAIVRGAVSSRKLAFLEFTGQ 3560
Qy 549 GAQTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPREVMAABEASVDAALLDQTAFTQ 608
Db 3561 GAQPMGMGRGLYETPAPAREAFDRCVALFDREIDQPLREVMAAAGLAQARLDQTAFAQ 3620
Qy 609 PALTFEEALAAALRSWGEPELVAGHSIGELVAACVAGVPSLEDAVFLVAARGRLMAL 668
Db 3621 PALPALFETALAAALRSWGEPELVAGHSIGELVAACVAGVPSLEDAVFLVAARGRLMAL 3680
Qy 669 PAGGAMVSIABEADVAALVAAPHAASVSIAAVNAPOVVIAGAGOPVHAIAAAMAARGAR 728
Db 3681 PAGGAMVIAIASEVAEVAASVAPHAATVSIIAVNGDAVVIAGAEVQVLAIGATPAARGIR 3740
Qy 729 TKALVSHAFHSPLMAPMLEAFGRVAESVSYRPSIVLSNLGKACTDEVSSPGYVVRH 788
Db 3741 TKRLAVSHAFHSPLMDPMLIEDFORVAATIAVARAPRPVSNVTVGHVAPETATPEYVVRH 3800
Qy 789 AREVVRPADYKALHAAGAGTFVEVGRKSTLLGLVACMPDAPRALLASSAGDERPATV 848
Db 3801 VRSVVRFGDGKALHAAGATFVEVGRPVLLGLIPACLGADAVLVESLIRADRSSECEV 3860
Qy 849 LEALGGLNAVGLVSMAGLPPSGGRRVPLPTYPMOREBYMIDTKADA----- 896
Db 3861 LAALGAMVYAKWGALDMKGVFPDGGARRVALPMTPMOREHNMIDLTPRSAAPAGIAGRWELA 3920
Qy 897 ----- 896
Db 3921 GVGLCMRGVALHVLISIGPRHQPFLGDHLVFGKVVVPGAFHVAVLISIAERWPERATEL 3980
Qy 897 ----- 902
Db 3981 TGVEFLKAIAMEPDQVEVLAHVLTPBAAGDGVLELATLAAETERRWTTHARGVPTD 4040
Qy 903 -APGA----- 906
Db 4041 GAFGALPRELVELEBRATQPLDPFAGFLDLASVRIGWGLMRLWLDGVRVGDASLALTIVPT 4100
Qy 907 ---GHD-----EVEGGA-----VR-GGD 921
Db 4101 YPNADHVAFLPHILLDNGFAVSLSTRSEPDGDPPLPFAVERVMMARVAGVRGCGV 4160
Qy 922 RRSa-----RLDHP-- 932
Db 4161 PRSQAFGVSVFLVDTEGEVVAEVEGFCRABPREVFLROESGASTALYRLDWPBEALP 4220
Qy 933 -----BSGR 936
Db 4221 DAPARIESWVVAAPGSEMAALATRLNRCVLAEPKLEALAGVSPAGVITCMEGGA 4280
Qy 937 REKVEAAGR----- 946
Db 4281 HEEAPAAAGVATBGLSVQALRDBRAVRLMVVTMGAVAVEAGERVQVATAPWGLGRTVM 4340
Qy 947 ----- 946
Db 4341 QERPELSCTLVLEBEADDAARSADVLLRELGRADDETQVAFSGRRVARLVKATPEGL 4400

```

```

Qy 947 -----PFRLEIDEPVLDHLVLRVTERRAFGLEVEIANDAGLSFNDVQALGWNPD 1000
Db 4401 LVPDASEVRLAEAGCKGLTDQRLAPQARRAPGEVEIKVTASGLNFRVTLAVLGM----- 4456
Qy 1001 LRGKRNPELLIGECACAGIYAVGCVNGLVYQGVYIALSAGAFTHYTTSAALVLRPQA 1060
Db 4457 YPGDAGR---MGGCACAVATVAGGVVHVGDAVMTL---GTLHRFTVVDRLVVRQPA 4511
Qy 1061 LSAIEAAMPVAYLTAVYALDRIRARLQGERVLIHAATGGVGLAAVQAQHVAGVACHTA 1120
Db 4512 LTPAQAATVPVAFITAMALHDGLNLRGERVLIHAAGAAGVGMVAQIARNTGAEVFA 4571
Qy 1121 GTPKRAVYLSLGY--RVVSDRSRDFVADVRAWTGEGVDVVLNSLSELIDXSFNLLR 1178
Db 4572 -SPSKMAAVQMGVPRTHIASRRLTEFAETFRQVTGGRGVDVVLNALAGEFVDASLSLS 4630
Qy 1179 SHGRFVELKRDCAVDNQ--LGLRPFLRNTSFSVLVDLRGMMLERPARVALLBELGLIA 1236
Db 4631 TGGRFLEMGKTD--TRDBAAVAAAHGVRVYVFDILEL-----APDRTRILERIVEGFA 4683
Qy 1237 AGVFTPPRIATLPIARVADAFRSMAQONHGXVLTLDGDEVOQIRITP-----HAGAP- 1290
Db 4684 AGHURALPVHAFATKAEAFREMAQRHOGKVL--LPASAPLAPFTGVLLTGGLGAL 4742
Qy 1291 -----STGDRDLDRILASAPARAALAEFLRTQVSQVLRTPETIKVG 1333
Db 4743 GILHARVLAQGVPHMVLTRRG--LDTPGAAKAVAEIEALGARITIASDVADNNALEAV 4801
Qy 1334 AEAL-----FTRLCMDSIMAVELNRRLEASLKLKIST 1365
Db 4802 LQAIAPAEWPLQGVYHAAGALDDGVLDGQTTDRFSFRVLAPKVTGAMNHLELTAGNDLAFV 4861
Qy 1366 TELSTS-----PNTALLAQNLDAATLASERVAENLRAGVOND 1406
Db 4862 LFSSMGLLSAGGSNTYA--ANNTFLDLAAHRAEGLAQSILANGPMSD 4909

RESULT 4
AAB83974
ID AAB83974 standard; Protein: 2153 AA.
XX
XX AAB83974;
XX
DT 06-AUG-2001 (first entry)
XX
DE Amino acid sequence of a type I polyketide synthase.
XX
XX Metabolic pathway operon; polyketide; polyketide antibiotic;
XX type I polyketide synthase.
XX
XX OS unidentified.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 1
XX FT /note= "Met encoded by GTG"
XX
XX PN W0200140497-A2.
XX
XX PD 07-JUN-2001.
XX
XX PF 27-NOV-2000; 2000WO-FR03311.
XX
XX PR 29-NOV-1999; 99FR-0015032.
XX PR 07-JUN-2000; 2000US-0209800.
XX
XX PA (AVET ) AVENTIS PHARMA SA.
XX
XX PI Jeanm P. Pernodet J. Guerin M. Simonet P. Courtois S.
XX PI Cappellano C. Francou F. Raynal A. Ball M. Sezonov G. Tuphile K.
XX PI Firostegard A.
XX
XX DR WPI; 2001-374849/39.

```





QY 1276 ----PEVO-----IRIPTHAGA----- 1288  
|||  
DB 1801 RASPEVQQAIAVMDADVTVQADVQSDLELERSVSSIDRLRGVTHAAVLDALLNQT 1860  
QY 1289 ----GPSTGD----- 1294  
DB 1861 EAHFRRWAAKIDGAMNHLTLTRDCPLDHFVLFSSAGLGLGPAQGNVAAANAFDLALAY 1920  
QY 1295 ----RDLD----- 1299  
DB 1921 YRKAQGLPALSIGWAMSEVGLAAQDNRSRLALRGMENTLPQGLALILEQLNNSACH 1980  
QY 1300 ----RLASAPARAALAEAF 1316  
DB 1981 VAAMPINVRQKQFYPKAAQSALPELHDDAASDAPNALRAHQSAEPQTRTLEEH 2040  
QY 1317 LRTQVSQVLTPEIKVGAELFTRLGMSLMAVELNRNIEASLKIKSTTFLSTSPNIAL 1376  
DB 2041 LQQDLARVLRIDSCITDRLPELKEGSPSLMALFERNRLLETLGLTPATLWGHPTLAG 2100  
QY 1377 LAQNLDLATLALSLERFAA 1396  
DB 2101 LAPHLASQWGLPLVEAQAAA 2120

## RESULT 5

ABG99871  
ID ABG99871 standard; Protein; 4038 AA.

AC ABG99871;

DT 16-JAN-2003 (first entry)

DE S. cinnamomensis MonAIV/polyketide synthase multi-enzyme MONS4.

XX Monensin; gene cluster; polyketide synthase; antibiotic;

KM antihelminthic; insecticide; immunosuppressant; antifungal;

KW antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;

KV mon RI; mon RII; mon T; mon AIX; mon AX.

XX Streptomyces cinnamomensis.

OS MO200168867-A1.

PN 20-SEP-2001.

PD 30-MAY-2000; 2000WO-GB02072.

PF 28-MAY-1999; 99GB-0012563.

PR (BIOT-) BIOTICA TECHNOLOGY LTD.

PI Leadlay PF, Staunton J, Olynyk M;

XX WPI, 2001-611393/70.

DR N-PSDB; ABX04971.

XX New DNA sequence encoding polyketide synthase, useful for the

PT production of polyketides such as antibiotic monensin.

XX Claim 6; Page 82-85; 212pp; English.

XX The invention relates to a DNA sequence which is a fully defined sequence

CC of 103551 base pairs appearing as ABX04971, or its variant, that it is

CC not a sequence encoding all or part amino acids 1-920 encoded by mon A1

CC as given in the specification. The DNA is the S. cinnamomensis

CC polyketide antibiotic monensin biosynthetic gene cluster. Also included

CC are a recombinant cloning or expression vector comprising the gene

CC cluster, a transformed host cell which has been transformed to

CC contain the gene cluster (and is capable of expressing a corresponding

CC polypeptide), a hybridization probe derived from the gene cluster

CC (for identification and isolation of the same or analogous gene cluster,

CC e.g. one which binds specifically to a region of the monensin gene

CC cluster selected from mon BI, mon BII, mon CI, mon CII, mon H, Mon RI,

CC mon RII, mon T, mon AIX and mon AX), the use of the mon RI gene or

CC variant and a monensin promoter to control expression of a heterologous

CC gene in Streptomyces cinnamomensis, a polypeptide encoded by a portion of

CC the monensin gene cluster (preferably comprising mon BI, mon BII,

CC mon AIX or mon AX or their mutants, alleles or variants), an epoxide

CC enzyme encoded by mon CI, a cyclase enzyme encoded by mon CII, producing

CC S. cinnamomensis capable of enhanced levels of production of monensin

CC comprising engineering it to overexpress the mon RI gene, S.

CC cinnamomensis containing multiple copies of the mon RI gene and/or its

CC variants, expressing a gene heterologous to S. cinnamomensis comprising

CC transforming S. cinnamomensis with DNA encoding a heterologous gene and

CC expressing the gene under control of the activator gene mon RI or

CC actII/orf4 and 13-propyl erythromycin A. The processes and materials

CC (enzyme systems, nucleic acids and vectors) are useful for preparing

CC polyketides by recombinant synthesis. The polyketides are useful as

CC insecticides, antibiotics, antihelmintics, antifungals, antibacterials or

CC other pharmaceuticals. In particular the gene is useful for the

CC production of monensin, an antibiotic polyether polyketide. The

CC present sequence represents a protein encoded by the monensin gene

CC cluster.

XX Sequence 4038 AA;

Query Match 32.8%; Score 2363; DB 22; Length 4038;

Best Local Similarity 33.6%; Pred. No. 8,8e-179;

Matches 647; Conservative 193; Mismatches 515; Indels 572; Gaps 47;

QY 2 ADRIERABDPAIVAGASCRLPGCVITDLSGFWTLEGRDTCGRVPAER-WDAANFDP 60

DB 1823 AGADVADADTDPPVAIVAMTCRFPFGVASPDLDMLDLARCKAMGAFPDRCMDERLHP 1882.

QY 61 DPDAKPTVTRASFLSDVACFDSPFGISPREALRMDPARLLLEYCEALENAAIAPS 120

DB 1883 DPDPHTSYTDQGGFLPDAGDFDAFPFGINREALAMPQOGLLESMVELEAGIDPT 1942

QY 121 ALVTEGVFICIGPSEYEALPQATASAEIDAHGIGTSPVAGASISVALGRPCVA 180

DB 1943 TLKGTPTGTVMYHYAKSFP--TADQOLEGVSYLASTSVMGSRVATLGEGPAVT 2000

QY 181 VDIAYSSSLVAHLACQSLNSGECSTLACGVSIMLSPSTLVLSKTRALARDGCAFS 240

DB 2001 VDIACSSSLVSIHLATQALRRGECDLALAGGVTMADPDMFAGSRQSGISPDRCRKA 2060

QY 241 AADGPRGECCAVVLKRLSGARADGRIIAVIRGSAIINDGASGGLTVPNGSOEIVL 300

DB 2061 AADGVPSEGVGLLRLSDARRHGRVYGVVRSAGAVNDGASNGLTAFNGPSQERVI 2120

QY 301 KRALADGCAASVGVYEAHGTGTTLDPIEIQALNAVYGIRDVATPPLIGSVKYNIGH 360

DB 2121 KQALASGGLSSVDVYVEGHGTGTTLDPIEIQALNAVYGGREDBRPLMGVSKNIGH 2180

QY 361 PEVASGITGLKLVLSLQHQIPALHQAQALNPISW--GDLRLTVTRATPMD-WNTP 417

DB 2181 TQAAAGVAGVIMYMARHGVVPASLHVDPSPHVEWDSCAVRLAV--ESVPQWVEGRP 2238

QY 418 RRAVSSFGMSGTAAHVLEFAP-----AATCTPPAPERAP-----LVASARTASLD 467

DB 2239 RRAVSSFGMSGTAAHVLEFAP-----AATCTPPAPERAP-----LVASARTASLD 467

QY 468 AQAARLDHLETVSQC-IGDVAFSLATTSAMEHRLVAATSEGLRAALDAACQOTS 526

DB 2299 DQALRLDFASDAFRAPIADVGSLLKTRALHHRVAVVGAEEAELALAEALATGEPH 2358

QY 527 PGAVRSIADSR---GKLAFLTGQAGTQGLGGLYDVMSAFREAPDLCVRLFNQELDR 583

DB 2359 AALVGPAQSGARVGGDDVVMFLFSGQSOLVGMAGLVERFPVFAAEFEVCGL----LEG 2414

QY 584 P-----LRVMAEAPASVDALLDOTAFQPLLFTEVYALALWPSWGPBPLVAGHS 636

DB 2415 PLVGAAGGLREVVRGPRE---RLDHTVMAQGLFALQVGLARLMESVGVPRPDVLAGHS 2470

```

QY 637 IGEIWAACVAGVFESEDAVFLVAARGRLMQLPAGAMVSTEAPDAVAAPHAASV 696
DB 2471 IGEIAAAHVAGFDLADACRVGAPRLMGLPEGGAKCAVOATPELADV--DGSAAV 2528
QY 697 IAAVAPQOVNLAGAGQPVHAIAMAAAGARTKALHVSHPSPMLAPMBAFGVABS 756
DB 2529 VAAVTPSTVTSIGFSDVDRLAGVWRERGRKTKLSVSHAFSHALMEFMLEFTEALRG 2588
QY 757 VSYRPSIVLVNLSGKACTDEVSSFGYVWRHAREVVRPADGVKALHAAAGCTFVEVGPK 816
DB 2589 VYFRPSPISIPMSNVSGERAGEEITDPEYWRHVRNVAFLFOPAI-AQVADSAVFEVLGPA 2647
QY 817 STLGLGVPCW--PARAPALLASRRAG--RDEPATVLEALGMAVAGLVSMAGLPSGGR 873
DB 2648 PVLTTAAQHTLDBSDSQSVLASLAGEEPESAFVEAMARLHTGAVADMVSLF--AGD 2705
QY 874 RVP-----LPTYPWORERYWIDTK----- 892
DB 2706 RVPGLVELPTVAFQHERFWLSGRSGGDAATLGLVAAGHPLLGAAVFADRGCLLTGRL 2765
QY 893 -----ADDAARDDRAPGA-----GHDE----- 910
DB 2766 SRSGVSWLADHVAGAVLVPAGALVEMALRAGDEVGCTVEELMLQAPLVEASGLRVQ 2825
QY 911 --VEBGG--AVRGDR--RSARLD--HPP-- 931
DB 2826 VVEEAGDEDRRGVOIYSRPDADAVGGDSWICHATGLVSPESARLDTGLGVPAPAGAE 2885
QY 932 -----PES----- 934
DB 2886 PLVDGFFVAQAGEAGYGPAPRGLRAVVRHGODLLAEVVLPEAGAHNDGYGIHPALLDA 2945
QY 935 ----- 934
DB 2946 TLHPLLAARFMDSEDDQLYVFGWAGVSLRAVGATTVVRRLPVGESVDCGLSVTVTDA 3005
QY 935 -----GR 936
DB 3006 TCGPVLVSUDSLQTRPVKESQLAAQOPDVRCGLFTVEMTPLPOTDADGEADMVVLSDGVCR 3065
QY 937 -REKTEAAG----- 944
DB 3066 LADVVAAGGEAPMAVAVPVDASVGDGRGRLDRLVVERVLSVQEFLLPELASRLIV 3125
QY 945 -----DRPRL----- 950
DB 3126 VTRGAVATGVDGDGDVDAASAAVWGLVRSAGSENPGRFILLDVGDDGDQDGLNGRLHP 3185
QY 951 -----EIDEP-----GVLDHLVLKTERR-----APG----- 972
DB 3186 HATLHAAEELDEFPOLAREGTL--VYRHLTQARQSAELVPPGEPAMLRVNHDSLDA 3243
QY 973 -----LGEVEIAVDAAGLSFNDVQLAGVPPDDLPGKPNPPLLIGECACGR 3295
DB 3244 LAAVACPBEALBPALPQGVRIAAHAGINFRDVLVNLGMP-----AYGAMGEGAGV 3295
QY 1019 IYAVGEGVNGLVVGCOPYVILSAGAFTHYTTSAALVLPPOALSIAEAAVAYITAMY 1078
DB 3296 VTEVSPREVTHVSVGRVWGVFEGAFPPVIAARVWTVPVQDMRMEAGIRPAFLTAMY 3355
QY 1079 ALDRITARLOPGERVLIHAATGVGLAAVQMAOHVGAENATGTPEKRAYLSLGV--RY 1136
DB 3356 GLVEILAGLAGEERVLHAATGVGMAVQIARHVGAEVATA-SPEKHAVLSEMGIDAAH 3414
QY 1137 VSDSRDFVADVAVMTGEGGVVNLNSGELIDKSPMLRSHGFVELGRDCYADNQ 1196
DB 3415 RASSDILAFEGTFRERATGGRMDVVLNSLAGEFIDSLFLDLDDGGRFLMGKTDPRAAE 3474
QY 1197 LGLRPFLRLSLVDLRGMMLERPARVALLBEILGLIAAGVTPPPIATLPIARVADA 1256
DB 3475 VAAEH--ADVSTAYDLVD--AGPDRISNMLDKVELLEASERLKLPLVRSWPLDKACEA 3530
QY 1257 FRSMQAQHLGKLVLTG--DPEVQIRIPTHAGAGPST-----GDRDL--DRLA 1302

```

```

DB 3531 FRFMSQAHTGKLVLEIPEALDPDECTVIVTGTGALGVVAEHLVRENGVNHLLASRG 3590
QY 1303 SAAPARAALBAFLRTGVSVGLRTPEIKVGAELFTRLGM-DSIMAVELNRLEASLKL 1361
DB 3591 PEAPG-----SDELASKLTGLGAEVTIVAADVSDPASVELVGKTPSHPL 3636
QY 1362 KLSTFPLSTSPRIALLAQNLLDALATALS---LERV-----AAENL-----RAGVQN 1405
DB 3637 -----TGVAHAAGVLEDDGVVTAQTPGLARVMAKAAANLHEATREKSLGLFV 3686
QY 1406 DFVSSGA 1412
DB 3687 VFSSAAA 3693

RESULT 6
AAE10145
ID AAE10145 standard; Protein; 5435 AA.
XX
AC AAE10145;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin gene, NysJ.
XX
KW polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic.
XX
OS Streptomyces noursei.
XX
PN W0200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UNO-) UNIV NORGES TEKNIISK NATURVITENSKAPELIGE.
PA (SNPE) SINTEF STITFELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FUAE/) FUAEVNIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR MPI: 2001-557614/62.
DR N-PSDB: AAD17186.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 15; Page 260-263; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a PKS encoding Streptomyces noursei nystatin gene, NysJ.
XX
SQ Sequence 5435 AA;
XX
Query Match 32.7%; Score 2355.5; DB 22; Length 5435;
Best Local Similarity 34.2%; Pred. No. 5.6e-178;

```

Matches 655; Conservative 183; Mismatches 513; Indels 563; Gaps 47;

```

Qy 3 DRPIRAEDPIAIAGASCRLPGVYIDLSGFWTLLGSRDITGVRPAER-MDAAMFPDP 61
Db 32 NRMVAAAKEPIAVVGMACRPFQAVDSPEALMEMVATGTVISGPPDRGMDLELRSG 91
Qy 62 PDAPKTEVTRASFSLDVACFDASFFGISPREALMDPAHRLLEVCWEALEMAIASA 121
Db 92 TDAROTVDSQSGFIDCIADFPFGFISPREAVTMDPQORLLTTAMEAVERAGIDATY 151
Qy 122 LVGTETGFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGRLPCAV 181
Db 152 LRAFTGAFIGNGDYVYLLVRSLDADTGV--GTGIAASAASGRSLYTLGLEPPALTV 209
Qy 182 PTAVSSSLVAVHALACOSLRSGECSTALAGVSLMSPSTLVMLSTRALARDGRCAKSA 241
Db 210 DTACSSSLVHLHLAVQALRNGECGMALGCVVMAVTPSGLVEFSQGLARDGRKAKAD 269
Qy 242 EADGFGEGCAVVYLLKRLSGARADGDRILAVIRGSAINHDGASGLTVPNSSQEIYVK 301
Db 270 AADGTGMEGAGVLLERLSDAQRNGHPVLAVVRSANVQDASNGFTAPNGPSQQRVIR 329
Qy 302 RALADAGCAASSVGVYEAHGTGTIGDPIETIOALNAVGLGRDVTPLIGSVKTNLGH 361
Db 330 QALNAGLATGDIIDAVEAHGTGTPLGDPTEAQSILATY--GQDRAHAPVLLSISKNMHT 387
Qy 362 EYASGITGLKLVLSLQHQIPAHILHAQALNPRISM--GDRLITVTRATPMDNTPRR 419
Db 388 QASAVVAGVITKIMAMRHGVLPRTLHVDRPSTHVMTTGSVEL-LTDH-PMPEGRRR 445
Qy 420 AGVSSFGMSGTNAHVLEEAR-----AATCTPAPERPAELVLISARTASALDAQARLR 474
Db 446 TGISFGVSGTNAHYIVEQAPDTPAEADDPTRPRTLPM-LTARFAGALRDQATALL 504
Qy 475 DHLETPBQCIG---DVAFSLATRSAMERHLAAATSRGL--RALDAAACQTSFG 528
Db 505 DHLND-PGDRGPALDPAFSLATRALRHLAV-VTGTDTAGRDLLTMLAHGTAD 562
Qy 529 AVRSIADSSRGKLAFLFTGCGAQTGMRGLVDVMSAFREAFDLVRLFNOLDRPLAEV 588
Db 553 AHEGHA-AGRTRCALFSGQAGNORLGMGRHLHARPPVARALDTVLDLDELGGTLEEV 621
Qy 589 MAAEPASVDAALLDOTAFTOPALFTFEVALAALMRSWGPBELVAGHSIGELVACVAV 648
Db 622 IW----GTDDAPLNTGFTQPALFAVEVALYRLISWGAAPDFVAGHSIGELAAHVAVG 677
Qy 649 FSLERAVLVVARGKLMQALPAGGAMVSIPEARADVAAAVAPHAASVSIAAVAPDQYI 708
Db 678 FSLBEDACTLVAPRAGIMQALPRGAMVAVVATEDEVSLITD--GVAIAAINGPTSLV 734
Qy 709 AGAGCPVAHIAAAMARGARTKALHVSHPHSLMAAPMLAEAGRAVESVSRPSIVLS 768
Db 735 SGDEPATAVAAARLEOGRRTLRVSHAFHSLMDPMLAEFRAVAEGSLSGEPIPVAS 794
Qy 769 NLSGKACTDE-VSSPGYVWRHAREVFPADGVKALHAAGAGFVEVGEKSTLLAGVPACM 827
Db 795 NLTGAVADGTLLGTADYVWRHAREVFPADGIRALTDAGVGAFLGSDGTALAAQOSA 854
Qy 828 PDA--RPLLASSRAGDEPATVLEALGMAVGLVSWAGLFP--SGGRVPLPTYPWQR 884
Db 855 PDAVSVPVL---RKDRDEEPAVAAALARLHTAGVPVMTAFYAGTGARHTDLPYAAQY 910
Qy 885 ERYWID-----TKADDAARD----- 900
Db 911 ERYWKATYRPADATGLTLAADHPLLGAAVSVASDELTLTGLSLATHPMLADHVVG 970
Qy 901 -----RRAPGAGHDEVEE----- 913
Db 971 MVFFPGTFLLEAVAAADQGVCDRVEELMLAFLILPATGVQMQIANGAADDDGGRDLR 1030
Qy 914 -----G 914
Db 1031 FTRBDDPDAAMQAHATGRITGERVIALDTTTPPRDAEPVIDIDGLYRANGLDYG 1090

```

```

Qy 915 GAVRG-----GDRS--- 924
Db 1091 PVFRGLAAVRMRDREIYAVALPEGTADADAFGLHPLPDAVLHSTLPASADGDRGLP 1150
Qy 925 -----ARLDH 929
Db 1151 FANNVSLHAAGADALRVRITSCGPAVEITAVDPQGRPVVSVESLTLRAAGPDAGTADH 1210
Qy 930 -----PPESGRREVEAG- 944
Db 1211 RADAGSLFRMDWTRTYHAPATPATMAVLGTDPILGTLEALTAAQPDVTYGLDGVDLGE 1270
Qy 945 -----DRPE----- 948
Db 1271 LTAGDDRPVPVAVPLRGATDHGPAGAHDTTRTVLLLOEWLAERRPARSRLLVTRGA 1330
Qy 949 -----RLSIDP-- 955
Db 1331 VADGERGPLDAAAPVWGLVRSAGSENPGRLLLVLDLDTAESAAQLPLLALDLADEPQA 1390
Qy 956 -----GVLDHL-VLRATERAPGLG--EVE 977
Db 1391 VREGTVRVGRLARLDSGRGLVPPPGTPWRIGSRKAGSLDGLALLPPEARPLTGEHVR 1450
Qy 978 IAVDAAGLSPNDVOLALGMVDDLPGRKNPPLILGCECAGRIYAVGEGVGLVVGOPYIA 1037
Db 1451 VGIIRAGLNPFDVNALGMPDAG-----LFGSEAGVYVEGVEVIGLARGDRMVG 1503
Qy 1038 LSAQAFATHVTTSAALVPRPQALSAIEAAMPVAVYLTAMVALDRARLARGERYLTHAA 1097
Db 1504 MLFGGFGPLGIADARLLTPVPADWSMETGASVPLVPLTAYVALKELGRLAGEKYLVIAG 1563
Qy 1098 TGVGVLAVQWACHVGVAVHATAGTPERKATLESLGR--VYDSRSRPFADVRAWTGG 1155
Db 1564 AGGVGMAIQTARRVGAVEFATA-SEGRWDVLRSLGVAADHDIASSRTLDFEPAAREVAGD 1622
Qy 1156 EGVVDVNLNSLSEGLIDFSFNILRSGRFVELGKSDCYADNQLGRPLRNLSPSLVDLUG 1215
Db 1633 RGLDVLNALSGEYVDASMRLLSGGRFLGKTDIDAAADVP-----DGLSYHSFDL-G 1676
Qy 1216 MLEERPARVRLLELLGLIAAGVFTPPPIATLPIARVADAFRSMAQOHLGKVLTLG- 1274
Db 1677 MV--DPEHIGMLDLVLELFDRLGALALPVRSWDPVRAGEAFRMSLAQHGKIVLTVPQ 1734
Qy 1275 --DPEVQIRIFTAG--AG-----PSTGRDL--DLASAPARA----- 1310
Db 1735 PLDDPGTVLLTGTGGLARHLVTEHGRHLLLAGRRGPDAPGAALHAELTALGAE 1794
Qy 1311 ---AALEAFRTQVSOVLRTPERK-----VGAEALFTRLGMDSLMAVELRNRIE 1356
Db 1795 VTYAACVDADRITALAALATYPAEHPLTAVVHTAGVLDGDTLTLNPDRLATV-LRFPVD 1853
Qy 1357 AS-----LKLKSLTFPLSTS-----PNIALLA--QNLDLALAT---ALSL 1391
Db 1854 AAMHLHDLTRLLDLAAFLVYSSTGAVGPGQANVYAAGNTFLDLAAHRAHLGL 1907

```

RESULT 7  
ABG99872  
ID ABG99872 standard; Protein; 4106 AA.

ABG99872;  
16-JAN-2003 (first entry)

S. cinamomensis MonAv/polyketide synthase multi-enzyme MONS5.  
Monensin; gene cluster; polyketide synthase; antibiotic;  
antibacterial; insecticide; immunosuppressant; antifungal;  
antibacterial; polyether; mon B1; mon B11; mon C1; mon C11; mon H;  
mon R1; mon R11; mon T; mon AIX; mon AX.



```

Db      3092 VLSDGVSLADVNSAAGGEAPWAVAPVAGSAGGAGIAGDRREGIDGRLVVERVSLVOE 3151
Qy      923 -----RSARLDHP----- 930
Db      3152 FLABELAESRLVLTIRGAVATGDDGDVDASAAAWGLVKSASQENPGRFILLDVMND 3211
Qy      931 -----PPESGRR----- 937
Db      3212 VDVDVMDVDVDVDVDVDGNGSDLDLPDLNGRLPHATIRHAAEELDEPQLARDDQ 3271
Qy      938 -----EKVEBAG-----DRPFRL-----IDEPGLDHLVLRVTERARGL 973
Db      3272 LLVPLRVATGGGLVVAFTDRAMRLDKGSAETLESVAPVAPYGVMEPL-----GP 3321
Qy      974 GEVEIAVDAAGLSPFDVQALGMPVDDLPKGNPPLLGBEGAGRIIVAGEGVNGLVVQ 1033
Db      3322 GQVRIGIHAGINFEVDVLSLGMVPGVG-----LGBGAGVVTETGPDVTHLSVD 3373
Qy      1034 PVIALSAGAFATHTVTSALVLPRLPQALSAIBAAMPVAYLTAWYALDRIRARLQPERVL 1093
Db      3374 RVMGVLHGSFEPPTAVADRIMVAPVPGQMDMRQAAMPVAYLTAWYGLVELAGLKAGERVL 3433
Qy      1094 IHAATGVGLAAVQVAQVGAEVHAITAGTPEKRAVLESIGV--RYVSDRSRDPVADVRA 1151
Db      3434 IHAATGVGMAAVQIARHLAGAIVFATA--SAAKHVLTLEMGIDAARASRDLAFEDTRQ 3492
Qy      1152 WTGGGVUVVUNSLSGELIDKSFNLRSRGRFVELGKDCVANDQLGR--PFLRNLSTSL 1210
Db      3493 ATDGGGMVDVUNSLTGEFLDASLRLLGDDGRLEKTDVTPTEVVAEYPGVTVTVVDL 3552
Qy      1211 VDLRGMLEPRPARVRLLEELGLIAGVFTEPPITLPIARVADAFRSMAQOHLGKLV 1270
Db      3553 VTLDAG-----PDIRAVMSELGERFASGALDLPVRSWPLDKARAFRMSQAKRTGTVL 3607
Qy      1271 LTLG---DPEVOIRIPTHAGA-----GPSTGDDLLDR 1300
Db      3608 LDVPAPLPDGTVLITGTGALGQVVAEHLVREMGVRHLLASRRGLDAPGSG--ELADR 3665
Qy      1301 LASAPARAAALE 1314
Db      3666 LSDLAGAVTVAAAD 3679

```

```

RESULT 8
AAE10129
ID AAE10129 standard; Protein; 11096 AA.
AC AAE10129;
XX
XX
XX 29-NOV-2001 (first entry)
XX
XX
XX Streptomyces noursei nystatin gene, NysC.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; PKS type I.
XX
OS Streptomyces noursei.
XX
XX Key
XX FH Location/Qualifiers
XX FT 35..455
XX FT /label= KS3 domain
XX FT /note= "Ketosynthase (KS) domain"
XX FT 546..858
XX FT /label= AT3 domain
XX FT /note= "Acyltransferase (AT) domain"
XX FT 872..1073
XX FT /label= DH3 domain
XX FT /note= "Dehydratase (DH) domain"
XX FT 1381..1628
XX FT /label= KR3 domain
XX FT /note= "Ketoreductase (KR) domain"
XX FT 1662..1735
XX FT Domain

```

```

FT /label= ACP3 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 1757..2180
FT /label= KS4 domain
FT /note= "Ketosynthase (KS) domain"
FT 2291..2603
FT /label= AT4 domain
FT /note= "Acyltransferase (AT) domain"
FT 2617..2818
FT /label= DH4 domain
FT /note= "Dehydratase (DH) domain"
FT 3124..3371
FT /label= KR4 domain
FT /note= "Ketoreductase (KR) domain"
FT 3407..3480
FT /label= ACP4 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 3501..3924
FT /label= KS5 domain
FT /note= "Ketosynthase (KS) domain"
FT 4032..4346
FT /label= AT5 domain
FT /note= "Acyltransferase (AT) domain"
FT 4360..4561
FT /label= DH5 domain
FT /note= "Dehydratase (DH) domain"
FT 4953..5239
FT /label= ER5 domain
FT /note= "Enoylreductase (ER) domain"
FT 5248..5495
FT /label= KR5 domain
FT /note= "Ketoreductase (KR) domain"
FT 5528..5601
FT /label= ACP5 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 5623..6046
FT /label= KS6 domain
FT /note= "Ketosynthase (KS) domain"
FT 6165..6478
FT /label= AT6 domain
FT /note= "Acyltransferase (AT) domain"
FT 6492..6704
FT /label= DH6 domain
FT /note= "Dehydratase (DH) domain"
FT 7038..7281
FT /label= KR6 domain
FT /note= "Ketoreductase (KR) domain"
FT 7315..7388
FT /label= ACP6 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 7408..7831
FT /label= KS7 domain
FT /note= "Ketosynthase (KS) domain"
FT 7939..8253
FT /label= AT7 domain
FT /note= "Acyltransferase (AT) domain"
FT 8267..8470
FT /label= DH7 domain
FT /note= "Dehydratase (DH) domain"
FT 8812..9086
FT /label= KR7 domain
FT /note= "Ketoreductase (KR) domain"
FT 9120..9193
FT /label= ACP7 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 9214..9637
FT /label= KS8 domain
FT /note= "Ketosynthase (KS) domain"
FT 9758..10072
FT /label= AT8 domain
FT /note= "Acyltransferase (AT) domain"
FT 10086..10289
FT /label= DH8 domain

```

```

FT      /note="Dehydratase (DH) domain"
FT      Domain 10657..10904
FT      /label=KR8_domain
FT      /note="Ketoreductase (KR) domain"
FT      Domain 10939..11012
FT      /label=ACP8_domain
FT      /note="Acyl carrier protein (ACP) domain"
XX      WO200159126-A2.
XX      16-AUG-2001.
XX      08-FEB-2001; 2001WO-GB0509.
XX      08-FEB-2000; 2000GB-0002840.
XX      10-APR-2000; 2000GB-0008786.
XX      14-APR-2000; 2000GB-0009387.
XX      (UYNO-) UNIT NORGES TEKNIKSKAPETJELIGE.
XX      (SNTP) SINTER STIPELSEN IND TEK FORSK.
XX      (ALPH-) ALPHAKMA AS.
XX      (SINV-) SINEVENT AS.
XX      (DZIE/) DZIEGLEMSKA H.
XX      (ZOTC/) ZOTCHEV S B.
XX      (SEKU/) SEKUROVA O N.
XX      (FJAE/) FJAEVUK E.
XX      (BRAU/) BRAUTASET T.
XX      (STROM) STROM A R.
XX      Zolchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX      Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX      MPI; 2001-557614/62.
XX      N-PSDB; AAD17184.
XX      New nystatin polyketide synthase polynucleotides and polypeptides,
XX      useful as antibiotics and antifungals -
XX      Claim 15; Page 170-176; 266pp; English.
XX      The present invention relates to the cloning and sequencing of the gene
XX      cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX      involved in the biosynthesis of the macrocyclic antibiotic nystatin.
XX      The nystatin PKS is useful as antifungal antibiotics. The present
XX      sequence is a PKS type I encoding Streptomyces noursei nystatin
XX      gene, NysC.
XX      Sequence 11096 AA;
SQ      Query Match 32.3%; Score 2331.5; DB 22; Length 11096;
SQ      Best Local Similarity 32.9%; Pred. No. 1.4e-175;
SQ      Matches 629; Conservative 162; Mismatches 519; Indels 583; Gaps 34;
XX      10 AEDPIAVGASCRLPQGVLDLGSFWTLLEGRDVTGRVPAER-WDAAMFDPDPDAAGKT 68
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      3499 ADDPIYVGMACRFPGGVTSPEBDMRLVDDGDAITTFPTNRRMDLNDLYDPREHFGTS 3558
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      69 PYTRASFLSDVACFDASFFGQISPREALRMDPAHRLLEVCWEALENAALAPSAALVGTETG 128
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      3559 YTRSGGFLHEAGFEFDPAFGMSPREALATDSOQRLTLESSWEAIEBAGIDPLTRGSATG 3618
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      129 VPIGIGPSFEYEAALPATASAEIDAAGCIGTMSVAGRIAYLGLRGCVANDTVYSSS 188
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      3619 VFGVWYISYIGTL---GGEFEGFGQGSAGSVASGRVYALGEGPAVTVDTACSSS 3674
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      189 LVAVALACOSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARDGRCAFASEADGFG 248
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      3675 LVALHMAAQAALRAGECSLALAGVTWMTSTPSTFVERSGRLAPDRSRAFAAALAGVGM 3734
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      249 GEGCAVVLKRLSGAPADGRIILAVIRGSALNHGASSGLTVPNGSSQEIYLRALADAG 308
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX      3735 SEGVGILVIERQDAVRNGHEILAVIRGSALNHGASNGILTAVNPGSQOQVIRQALASGG 3794
XX      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

QY      309 CAASGVYEAGGTGTTGDDPIEIOALNAVYGLGRDVATPILLIGSVKTNLGHPEYASGIT 368
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      3795 ISTDADVAEAGTITTDPIEAOALATIGRDRDPENPLMGLSKNIGITQAAAGVA 3854
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      369 GLKLVLSLONGOIPAHLLHAQALNPRISGDLRLTVTRARTWPDMPNTPRRAGVSSFGMS 428
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      3855 GVIKVMAMRHGVLHQTLLHVDAPS SHVMSVGAVELLIEQIAMPETGVRRAVGSFGIS 3914
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      429 GTNAHVLE-----EAPATCTPPAPERPAEL-LVLSARTASALDQAARLDHLETPS 482
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      3915 GTNAHVIVEQPALVSSPA---EPGRBPGVVPLPLSKSPALDQAARLLAGLAERPA 3971
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      483 QCLGVASFATTRSGAMEHRLAVALTSREGLEAALDAQAQGTSPGARSINDSSRGKLA 542
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      3972 LRPDLGSLATTRKAPFHRHRAVLATDPAADVALTALLAADADLSAV--VDDTTGCHA 4029
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      543 FLFTQGAQTLGMRGLYDVMSAFREAPDLCYRLFNOEL--DRPLREVMMAEPASVDAAL 600
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4030 VLFSGQSGRLGMRGELVERFVFALDVAIDHLDALPAQASLREVMWGQ---DVEL 4085
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      601 LDQTAFTQPALFTFEYALALMRSGVEBELVAGHSIGELVAAACVAFSLDEAVFLVAA 660
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4086 LDETGMTQPALFAVVALFRLVESGVRPDPFVAGHSIGEIAAAHVGVFSLDACLRLVAA 4145
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      661 RGRIMQALPAGGAMVSEAPADVAAVAAPHAASVIAANVAPDOVVIAGAQOPHAATA 720
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4146 RATLMQALPTGAMATAIOAEDEVTOHUTD--DIVIAAVNGFTVVVSGAESARTYAD 4202
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      721 AMAARGARTKALVSHAFHSPMLAEMLEAFGRVAESVSYRPSIVLVNSLQKACT-DEV 779
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4203 RLAEGRKTRRLRVSHAFHSPMLAEMLEAFGRVAESVSYRPSIVLVNSLQKACT-DEV 4262
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      780 SSPGYVWHAARVAFADGVYKALHAAGACTPVEVPEKSTLGLVACMP-DARPAALLASS 838
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4263 CSAEYMAHVRBAVAFADGVSTLENEGVTTFLIEAPDGLVLAQAQSLTGD--ATVPAL 4320
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      839 RAGRDEPATVLEALGLMAVGLVSMAGLFP--SGGRVPLPTYPWQRRRYW----IDTK- 892
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4321 KRDREERSALTALHLHTAGLRVMAAFPAAGSAGRDLPTTYAQHATYWTGTLPLPAH 4380
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      893 -----ADDAARGRRAPG----- 905
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4381 AAAYGLTAAEHPLNGSVLEAGEGVLTGRLSLQSHPMLDHVMAGVLLPGTALLLELA 4440
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      906 -----AGHDEV-----EGAV----- 917
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      4441 FRAGDEAGCDRVEELTLAAPLVLPERGAVQTQVRGVADDTGRTVTVHSRBEHATDVSW 4500
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      918 -----R 918
QY      4501 TQHATGTLTMSGAPADTGFDATAMPAPDAEPLATDCCARFTTLGFAVGPVQGLRAAMR 4560
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      919 GGD----- 921
QY      4561 AGDVLYAEVLAESTGDEBATAFGHLPALDLDAALHASLVAHGESEBSNGCLPESWEGATLYA 4620
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      922 ----- 921
QY      4621 TGATATLRYRLPTGTGDSVALIADTAGRPVAILDNLVSRVSGDQLTGAAGLARDALF 4680
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      922 ----- 921
QY      4681 TLDMNPVENLVENPVENVTGGGHAQODGRPAATYVALVGADDTALTAADTLTAAGIHHT 4740
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      922 ----- 921
QY      4741 LHPDLTLTATDADVPTVLLPLTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      922 -----RBSAR----- 926
QY      4801 ISTAALALVQEWTAQERFAGSRFLAVTTGTATAGGTDVMDVAAAVWGLVNSAQSEAPDT 4860
QY      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      927 ---LDHPPESGRREKVENA----- 943

```

```

Db      4861 FVLIRDPGPAGTHRTAAERQQLLRALHTDEPOLALRDGVLAAARLAREDTAAALTP 4920
      944 -GDRPREIDEKGVLDHLVLR-VITERAPGIG-EVEIIVDAAGLSPNDVOLALMVPDD 1000
      4921 PADRRMRRLSTAKSGINGLALTPYPALALPLTGHREVRVRAAGLNFRLVALALMVPDD 4980
      1001 LFGKPNPPLLTGECGACRIIVAVEGVNGLVGPVTAISAGAFAFHTVTSALVLPPOA 1060
      4981 DVGSG-----FGSEAAGVVEVGPETVGLAPGDQVMGMTSGSGSLAVDADARLARLPED 5034
      1061 ISAIEAAMPVAYLTAMVALDRIARIQPERVLIHAATGSGVGLAAVQAQHVGAEVHATA 1120
      5035 WEMETGASVPLVFLFYALKEIGLRAGEKYLVAHAGAGVGAALQIARHGAIEVFATA 5094
      1121 GPPEKRAYLESIGVR--YVSDRSRDFVADVRAAMGCEVDVNLNLSGELLDKSFNLIR 1178
      5095 -SEGKMDVRSLSGVADDDHTASSRTLDFEAPAEVAGDRGLDVNLNSLADFPVDAISRLLG 5153
      1179 SHGRFVELGKRDQYADNQLRPLRLNLSFSLVDLRGMLEPARPARVALLLELLGLIAG 1238
      5154 DGRFLEMKKTDIRADSVF-----DGLSYQSFDLAMVY---PETIGTALALMDLFRFG 5205
      1239 VETPPPIATLPIARVADAFRMAQAQHLKVLTL---GDPEVOAIRPTHAGAG----- 1289
      5206 ALRPLPVRTWDRHAKDAFRFMSMAKHIGKIVTLPRSWKPEGTVLVT--GGTGGIGGLV 5263
      1290 -----PSTGDRDLRLRLAARAPARAALEFLRTQVSOVLATPEIKVAEALFTRLGM- 1343
      5264 ARLHVRSCGVRHLTLTSSRGVGAAGAAGLVAELES-----IGARVVVAACDVG 5311
      1344 DSLMAVELNRLEASLKLSTTFLSTSPNIALAQNLLDALATALSERVA 1396
      5312 DCSAAVELVAGVSESTPLSA-----VHAAGVLDGVTGLSPERLAA 5354

RESULT 9
ABP57679
ID ABP57679 standard; Protein; 2149 AA.
XX
AC ABP57679;
DT 22-JAN-2003 (first entry)
DE Saccharopolyspora busb butenyl-epinosyn biosynthetic gene product.
KM Butenyl, biosynthetic enzyme; PKS; polyketide synthetase; macroliide;
KW metabolite; spinosyn.
XX
OS Saccharopolyspora sp..
XX
PN WO200279477-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-US09968.
XX
PR 30-MAR-2001; 2001US-280175P.
XX
PA (DOWC) DOW AGROSCIENCES LLC.
XX
PI Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
PI Mitchell JC;
XX
DR WPI; 2003-058434/05.
DR N-PSDB; ABV75557.
XX
PT New butenyl-spinosyn biosynthetic genes, useful for increasing the
PT production of butenyl-spinosyn insecticidal macroliides, or for changing
PT the metabolites or products produced by spinosyn-producing
PT microorganisms.
XX
PS Claim 1; Page 134-141; 218pp; English.

```

```

XX
CC The invention relates to a novel DNA molecule comprising a DNA sequence
CC that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC PKS (polyketide synthetase) domain, or a spinosyn PKS module. The
CC butenyl-spinosyn biosynthetic genes are useful for increasing the
CC production of butenyl-spinosyn insecticidal macroliides. The genes are
CC also useful for changing the metabolites or products produced by
CC spinosyn-producing microorganisms. The present sequence represents one of
CC the butenyl-spinosyn biosynthetic polypeptides of the invention.
SQ
Sequence 2149 AA;
Query Match 32.2%; Score 219.5; DB 24; Length 2149;
Best Local Similarity 32.8%; Pred. No. 1e-175;
Matches 610; Conservative 198; Mismatches 547; Indels 507; Gaps 32;
9 AAE--DPIATVAGSRLFGVYIDISGFTLLGSRDTYGRPALER-WDAAMFPDDPAP 65
30 AAEKGDPIAIVAMSGRYPGVQVSSPEDLWQLAAGVDALSEVPGDGMGLAGFPDSDRP 89
66 GKTPTVTRASFSDVACPDASFEGISPREALAMPDPAHRLLLEVCMEALENAAIAPGALVGT 125
90 GTSTACAGGFLQGVSEDFAGFFGISPREALAMPDQRLLEVAWEVFERAGLEGSTGGS 149
126 ETGVFETIGSPSEYEALPQATASAEIDAAGLGITWPSVAGRISYALGLRGPVAVDTAY 185
150 RVGVFVGTNGGDYASWL--RTPSEVAGHVLTGGAAILSTRVAVSFGFEGPAVTVDTC 207
186 SSSLVAVHLACQSLRSGECSTALAGVSLMSPSTLVMLSTRLALRGRCKASAEADG 245
208 SSSLVAHLAAGQALRAGCEDLALAGVTVMTSTPKAFLEFFSRORGLAAGRCCKSFMAADG 267
246 FGRGEGCAVVVLKLSLRGADSDRIYVRSATINHODASSGLTPVNSOEIYVLRALA 305
268 TGMGEGAGLLERLSDRRNGHRVLAVRSAAVQDASNGLTAPNSSQARVITQALA 327
306 DAGCAASSVGYVEAHGTGTTGDDPIETIOALNAVVGIGRDVATPLLIGSKTNLGHPEYAS 365
328 SAGLSVSDVDAVEAHGTITRIGDPIEAQALATYGRDDBPARPLMIGSVKNSIGHTQAAA 387
366 GITGLKTVLSLQHQIPIAHLHAQALNPRISWGDLRLTVTRAPRTFPMQMTPRRAGVSF 425
388 GVAGVTKVMAMRRHQLPRTLHVDAFSPDEVMSAGTVQLITEMMLPESGRRRAGVSSF 447
426 GMSGNNAVVLBEAPAACTPAPER-----PAELVLSARTASALDAQARLDHET 479
448 GISGTNAHVILBQPTGETRQAGPDSGVVDVPPVPMVSGKTPDALGADTLMSYDD 507
480 YPSQCLGDAVAFSLATTSAMEHRLAVATSRREGRLAALDAAACQTSFGAVRSIADSSRG 539
508 RVDVPSLDIAVSLMTRFALDBRAVVLGPDRETLISGLKALSAGHEAGVVTGSGVTC-G 566
540 KLAFLFTGQAGQTLGMRGLYDVMSAFREAFD-LCVRL--FNQELDRPLREVMMAEPAS 595
567 RIGFVFSGGQGMIGMGRGLYRAFPVFPAADFEACAELEAHGQGVG--VRDVVFGS--- 621
596 VDAALDQTAFTQPLFTFEYFALAAWMSKVEPELVGHSIGELVAVCAVGVSELEADY 655
622 -DAQLLNFTLWAGSLFALQVGLKLEKLDLSWGRPSAVVGHVSGELAAAFAGVLSLSDPA 680
656 FLVARGRLMOALPAGAMVSIAPADVAVAVPAHASVSIANVANADQVVIAGOPV 715
681 RLVAGRALMOALPBGCGMLAAVAGEBQRLPLADHGRVGLGANVNAESVVLSSDRVL 740
716 HAIAAAMARGAARTKALHVSHPHSLPMAFPLAEAFGRVAESVSYRPSIVLVNLSGA- 774
741 DDIAGRLDGCGVTRTFLVSHAFHSYRMDPMLDEFAEIAFARADVXRCELPIVSTLTGLKD 800
775 CTDEVSSGVVVRHAREVVRADGVKALHAAGAGFVGVGKSTLGLVPAQCM-----PDA 830
801 DAGRMSGDVPWVRQVREVRPADGQALVEHDVAVTIGIPDAGSALITQECVAASDOSR 860
831 RPALIASRAGRDEPATVLEALGILMAVGGLVSNAGLFP-SGGRVRPLPTYPMQRRERYMI 889

```



```

Db      861 RYAAVPMRNRNDEAQNLTALAOVHVRGAVDMRSEFFACTGAKOVELPTVAFGRQRYVL 920
QY      890 DTKADDAARAGDRRA-----PGAGHDEVEEGG----- 915
Db      921 ESDSDGVTVGAGLAGAEHPLGAVVPAAGDEVLTGRISVGTHPWLAEHRVLGEVIVPG 980
QY      916 ----AVGGDR-----RSARLDRP-----PESGRREV----- 940
Db      981 TALLELTALHAGERLGCERVEBELTLEAPLVPERGAMOVOLRVGAPENSGRRPWLVSYPE 1040
QY      941 ----EAGD----- 945
Db      1041 GAADHDWTRHATGRLAPGGGAAGDLADMPAGALPVDDEFRDLAEHGLEVEPIFOGL 1100
QY      946 ---- 945
Db      1101 KAAMROGEVYAALPGTEDSGFVHPALLDAAHATAVRMDMGLPFWMEGVCIHAR 1160
QY      946 ----RP----- 947
Db      1161 AASALRVVPAAGDAKSLVCDGTGRPVISVDRLVRSAAAGRTGARROHRAIRLIRLG 1220
QY      948 ---- 947
Db      1221 WPTVOLPTSAQPPSCVLGTSEVSDMOVYPDLRSI/LTALLDAGAPPGVIAPTPGGQ 1280
QY      948 ---- 947
Db      1281 TADVRESTRHALLDVOGMLADRLNDSRLFLVTRGAVAVEPGEPTDLAQAALMGLIRST 1340
QY      948 ---- 947
Db      1341 QTEHPRFVLVDVAEPAQLLPALPGVLAGEPQALRRGGAHAPRIAGIGDDVLPEPDS 1400
QY      948 --FRLEIDBPGVLDHLVLRTERRAP--GLGEVEIADVDAAGSFNDVOLATGMVDDLP 1002
Db      1401 MGRMLEATSPGTLIDGAL--LDEPAATASIGDGOVRITAMEAAGVNFADALIALGMVYG-- 1456
QY      1003 GKRPPELLIGBECAGRIVAVGEGVNGLVVGOPIYALSAGAFATHTVTSAALVIRPDALS 1062
Db      1457 ----AASIGGEGAGVYVETGPGVTGLAPGDRVMGMI/PKAFGLAVADHRMVTIRIPAGWS 1511
QY      1063 ATEAAMPVAVYLTAWALRIARLOGBEVLJHAATGVGLAAGVOMACHGAEVHAHTACT 1122
Db      1512 FQOASVPIVFLTAIYALVDLAGLRGEBESLWHSAGAGVGMATYQARHLAGEVATASE 1571
QY      1123 PEKRAYLESIGVRYVSDSRDFVADVRAMTGGEGVDVVLNLSIGELIDKSFNLLRSHGR 1182
Db      1572 DKMQAV--ELTRERLASSRTCPDQKQFLGATGGRGVDDVVLNSIAGDFADASLRMLPRGR 1629
QY      1183 FVELGKRDCTADNOLG-LRPFLENLSFSLVDLRGMMLERPARYALLBELGLIAAGVTT 1241
Db      1630 FLELGKTDVRDPEVADHAPGVSYQAFDVE-----AGFORGEMLDELVELFEFGVLE 1683
QY      1242 PPPIATLPFARVADARSMAOHGLKVLTLTGDPVQIRIPFHAGAGP-----STGDRD 1296
Db      1684 PUPVTAMVDYQAPFARHLSQARHVKLVINP-----PAMDTAGVTLVTTGGTG-- 1732
QY      1297 LLDRLASAPAAAPAAALAEFLRTQVSQVLRTEIKVGAE--ALFTRIAMD-SIMAVELELN 1353
Db      1733 ----ALGAEVAVRHLYTEHGVNRLVLSRRGPAASGAELVQLTLYGAEVSLQACDVAD 1787
QY      1354 RIEBASIKLSTFTSTSPRIALLAQNLLDALATALSLERV-----AAENRFAGVQ 1404
Db      1788 R-ETLAKVLAGIPDEHTLTAVVAAGVLDGVAESITAOQLDHLVRPKYDGAARNLIELTA 1846
QY      1405 ND 1406
Db      1847 PD 1848

```

```

AAV39298
ID AAV39298 standard; Protein; 2152 AA.
XX
AC AAV39298;
XX
DT 01-DEC-1999 (first entry)
XX
DE SpnB a polyketide synthase.
XX
KW Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW microtides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW PKS; extender module; initiator module; acyl transferase domain; AT;
KW acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW insecticide.
XX
OS Saccharopolyspora spinosa.
XX
FH Key
FH Location/Qualifiers
FT Domain
FT /label= KS2
FT /note= "Beta-ketosynthase domain: part of
FT /note= "extender module 2"
FT Domain
FT 536..886
FT /label= AT2
FT /note= "Acyl transferase domain: part of extender
FT /note= "module 2"
FT Domain
FT 892..1077
FT /label= ACP2
FT /note= "Acyl carrier protein domain: part of extender
FT /note= "module 2"
FT Domain
FT 1338..1683
FT /label= ER2
FT /note= "Enoyl reductase domain: part of extender
FT /note= "module 2"
FT Domain
FT 1687..1866
FT /label= KR2
FT /note= "Beta-ketoreductase domain: part of extender
FT /note= "module 2"
FT Domain
FT 1955..2034
FT /label= ACP2
FT /note= "Acyl carrier protein domain: part of extender
FT /note= "module 2"
XX
PN WO9946387-A1.
XX
PD 16-SEP-1999.
XX
PF 16-FEB-1999; 99WO-US03212.
XX
PR 09-MAR-1998; 98US-0036987.
XX
PA (DOWC ) DOW AGROSCIENCES LLC.
PI Balazs RH, Broughton MC, Crawford KP, Madduri K, Merlo DU;
PI Tredeaway PJ, Turner JR, Waldron C;
DR WPI, 1999-551414/46.
DR N-PSDB; AA221501.
XX
PS Claim 1; Page 96-103; 190pp; English.
XX
CC This is the amino acid sequence of the product of the spnB gene. The
CC protein is involved in spinosyn biosynthesis. The spnB gene is one of 23
CC genes and open reading frames contained in an 80kb DNA sequence
CC AA221501. Spinosyns are insecticidal microtides which are useful for the
CC control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC occurs via stepwise condensation and modification of carboxylic acid
CC precursors generating a linear polyketide which is modified further. The
CC DNA sequence contains a central region of approximately 55kb which has

```

CC homology to the DNA encoding the polyketide synthases (PKS) of known  
CC macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with  
CC stop codons at the end of acyl carrier protein (ACP) domains. Together  
CC the PKS polypeptides (AAV39297-Y39301) of which this sequence is one,  
CC form a complex consisting of an initiator module, spnA, and several  
CC extender modules spnB-spnE. Each extender module adds a specific acetyl  
CC Co-A precursor to a growing polyketide chain, and modifies the beta-keto  
CC group in a specific manner. A module in a PKS polypeptide consists of  
CC several domains with specific functions. The initiator module has an  
CC acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain.  
CC The extender modules have the same domains plus a beta-ketosynthase (KS)  
CC domain and optionally a beta-ketoreductase domain, a dehydratase (DH)  
CC domain, and an enoyl reductase (ER) domain. The last extender module  
CC terminates with a thioester domain. The products of the genes present  
CC in the upstream region of the PKS genes have been assigned names spnF-spnS  
CC AAV39302-Y39315 and are responsible for different modifications in  
CC spinosyn biosynthesis. There are also two ORFs ORF15 and ORF16 present  
CC immediately upstream of spnS, producing polypeptides AAV39316-Y39317,  
CC and two ORFs ORF1 and ORF2 present downstream of the PKS region  
CC producing polypeptides AAV39318-Y39319. The genes are useful to improve  
CC yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,  
CC or interruption of steps in spinosyn biosynthesis. The modified  
CC spinosyns may be a new insect control agent or serve as substrates for  
CC further chemical modification and the creation of new semi-synthetic  
CC spinosyns. The genes are also useful to isolate similar sequences from S.  
CC spinosa or other species by hybridization.

SQ Sequence 2152 AA;

Query Match 31.5%; Score 2273; DB 20; Length 2152;  
Best Local Similarity 32.7%; Pred. No. 5,3e-172;  
Matches 600; Conservative 199; Mismatches 542; Indels 496; Gaps 32;

QY 11 BDPPIVAGASCRIPGVIDLSGFWTLTSSGSDPTGVRPAER-WDAAMPDPDPAPGTP 69  
DB 36 DDPPIVAMSKRYPGQVSSPEDLMQAAAGVDAISEVEDRGMWDGCVFVPSDSDPGTSY 95  
QY 70 VTRASFLSDVACFDASFGISPREALRMDPAHRLLEVCWEALENAAIPASLVGTGCV 129  
DB 96 ACAGGFLQGVSEFDDGFGISPREALMDPQGRLLLEVAWEVEFAEGEQRSTRSGV 155  
QY 130 FIGICPSEYEALPQATSAETIDAHGIGTSPVQAGRIAYALGKPCVAVDTAYSSSL 189  
DB 156 FVGITGQDYASWL--RTPEPAVAGHVLTCGAAAVVSGHVAVSFGEGPAVTVDTACSSSL 213  
QY 190 VAVHLACSLRSGECSTLAGVSLMLSPSTLVMISKTRALARDRCFAFSAEADGPRG 249  
DB 214 VALHLAAGALRAGECDLALAGVATWSTPKVLEFSRQGLAPDRCKSFAGADGTWVG 273  
QY 250 EGCAYVVLKRLSGARADGRI LAVIRGSAIINHAGASSGLTVNNGSSOEIVLKRALADAGC 309  
DB 274 ECGAGLLRLERLSDARNGHEVLA VVRGSAVNDGASNGITAPNGSSQORVITQALASGL 333  
QY 310 AASVGVYEAHGTGTTGDPIEIQALNVA VYGLGRDVA TPILLGIVKYLGHPEVASGITG 369  
DB 334 SVSDVADEAHGTGRLGDPTEAQLIATYGRDRDPGRPLMLGYSKNSINGHTQAAGVAG 393  
QY 370 LKTVLSTQHQOIPAHILAQALNPRIISMGDLRLYTRARTPPMDMTRRRAGVSSFGMSG 429  
DB 394 VIKMVMARHGQLPFTLVHSPSPSEVDWSAGTVQLTENTPMPRSGRVRVGVSSFGISG 453  
QY 430 TNAHVLEAPATCTPPAP-----ERRPAELIVSARTASALDQAARLRDHLTYSQ 483  
DB 454 TNAHVILQPPGVPSQASGPGSGVVDVVPVPMWISGKTPRLSLQATALMTYIDERDV 513  
QY 484 CLGDVAFSLATTRSA MEHRLAVATSRREGLRALDAAAGQTS PGAVNSIADSSRGLAF 543  
DB 514 SSLDGVSLATLRRSLDERAAVVLGSDRETLICGVVALSAGHSAAGSLVTVSGVAG-GRIGF 572  
QY 544 LFTGGAGCTIGWGRGLYDVWSAFREAPD-LCYRL---FNQELDRLEREMAEPASVDA 599  
DB 573 VFSGGGGWLGKGRGLYRAFPVFAAFDEACAEHLAHIGQETG--VREV---SGSDAQ 626

QY 600 LLDOTAPALPTFEVALAALMRSWGEPELVAGHSIGELVAAACVAGFSLDPAVLVA 659  
DB 627 LLDRTLMQSGSLFALQVCLLTLIDSGWRPESVVLGSHSGEILAAFAAGVSLSGAARLVA 686  
QY 660 ARGTLMLPARGGAMVSTSEAPDVAANA VAPHAASVSTIAANADQVVIAGAGPVHIA 719  
DB 687 GRALMQALPFGGGLAVPAEEELMWSLIDQGRVGLAANAAGSVVLSGDRVLDLA 746  
QY 720 AAMARGARTALVSHAFFSLPMLAEPMLAERVAESVSRPSIVLVSNSGKA-CTDE 778  
DB 747 GRIDGGRKSWLVKVSARFHSYRMDPMLAEFAELARTVDYRCCEVPIVSTLTGLDDAGR 806  
QY 779 VSSPGYWRHAREVVRPADGVKALHAAGAGTFVEVGPSTLLGLVPACMPDA---RPAL 834  
DB 807 MSGPDYWRQVREVPVRFADVQVALVEHDVATVEIGPFGALSLALIQECVAASDHAGRLSA 866  
QY 835 LASSRAGDEPATVLEALGIMAVGLVSAGLPP-SGGRVPLPTPMQERV----- 888  
DB 867 VPAMRRNDEAQKWTALAHVHVGAVDWMSFPAGTGAKQILEPTVAFORQRYWLPSD 926  
QY 889 ----- 888  
QY 927 SGDYTGAGLAGAEHPLLAGVVPVAGGDEVLTGRISVTRTHWLAEBHYLGEVYIAGTALL 986  
DB 889 -----IDTRADDAAR 898  
QY 897 EIALHAGERLGCERVEBELTLEAPLVLPERGAIQVLRGAPENSGRRPMLYSRPEGAAE 1046  
QY 899 GD-----RRAPAGH----- 908  
DB 1047 HDWTRHATGRILAPRGENAAGLDADWPAPGALPVLDDEFYDLAEIGLEYGPIFOGLKAAW 1106  
QY 909 -DEV-----EEG----- 915  
DB 1107 RQGBEVYAEALPGBTEBSGFGVHPALLDALAHATAVDMMDARLPQWEGVSLAKAAPA 1166  
QY 916 -----AVRGDRRSA-----RLDHP- 930  
DB 1167 LRVAVPAGDAKSLVCDGTGRPVISVDRVLVLSAARFGARQARQLRYLSWPTV 1226  
QY 931 -----PREGRRKV 940  
DB 1227 QLPISAPPSCVLLGTSEVSADIQVYDRLSLTAALDAGAPPGVVIAPTPGGGRKADV 1286  
QY 941 E----- 941  
DB 1287 RETRHALDVQWLSQRLNESRLLVYTCQAVAVEGEPVTDLAQALMGLSRSTQTEH 1346  
QY 942 -----AACD-----RPR 949  
DB 1347 PDRVVLVDPPEADLLPALPEGVLA CGEPQALARRGGAHAPRLAIGSDVDLPVYDDGTWR 1406  
QY 950 LEIDEPGLDHLVIRVTERRAP---GLGEVEIAVDAAGLSFNDVOLAGVPPDLLPGKPN 1006  
DB 1407 LEATRPGLSDIAL-VDEPTATLADGGEVARIARRAAGVNRDMLIALGMVPG----- 1458  
QY 1007 PELLIGECAGRIYAVEGCVGLVVGQVYIALSAGAFATHTTSAALVLPPEQALSATEA 1066  
DB 1459 -VASLGSAGACVVETGGVGLAPGDRVMKMTPKAGCPPLAVADHRMVTIRIPACMSFARA 1517  
QY 1067 AAMPVAYLTAVYALDRIARLQGERVLIHAATGVGVLAAVQAOHVGAEVATAGTEKR 1126  
DB 1518 ASVPIVFLTAYALVLDLAGLRGSSLVHSAAGCVGMAAIDLAHLGAEVATASBDKWO 1577  
QY 1127 AYLESGLRVYSDSRDRFVADVRA MTGGSGVDVYVLSLSELIDKSPNLRSHGRFVEL 1186  
DB 1578 AV--ELSRHILASRTCDPFOQFLGATGGRVDVVLNSLAGEPADASIRMLPRGGRFLEL 1635  
QY 1187 GKRPYADNONG-LRPFRLNLSFSLVDRGMLEPRPARVALLEBELGLIAGVFTPEPI 1245  
DB 1636 GKTVDRDVEVADAHAPGVSTQAFDTE-----AGPDRIGMELVELBEGRVLEPDPV 1689  
QY 1246 ATPLEIARVADAFRSMAQOHLGKLVLTLDGPEVOIRIPTHAGAP-----STGDRDLDR 1300

DB 1690 TAMDVAQAEALRHLSQARHVGLVITMP-----PMDAAGTVLVTGGTG----- 1734

QY 1301 LASAPAPAPAAALAEAFLEKTOVSQVLTPEIKVGAE--ALFTRLGMD-SIMAVELNRRIA 1357

DB 1735 -ALGAEVAHVLIERGVNMLVLSRRGPAASGAELVAQLTAGAEVSLQACDVADR-ET 1792

QY 1358 SLKTKSTFTLSTSPNIALLAQNLDAALATLSLEEV 1394

DB 1793 LAKVLASIPDEHPLTAHVAAVLDGVSSESLTVERL 1829

RESULT 11

AAB70966

ID AAB70966 standard; Protein; 2152 AA.

AC AAB70966;

DT 28-AUG-2001 (first entry)

DE S. spinosa protein fragment encoded by ORF19, SEQ ID 44.

XX Forosamine; trimethylrhamsone; polyketide synthase; biosynthesis;

XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;

XX macrolide; insecticidal; polyketide synthase.

OS Saccharopolyspora spinosa.

PN DE19957268-A1.

XX 08-MAR-2001.

PD 29-NOV-1999; 99DE-1057268.

PR 27-AUG-1999; 99DE-1040596.

XX (FARB ) BAYER AG.

PI Eberz G, Moehle V, Froede R, Velten R, Salas JA;

XX MPI; 2001-267102/28.

DR N-PSDB; AAF88336.

XX

PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for

PT recombinant production of insecticidal spinosyns and their derivatives

XX

XX Claim 57; Page 200-209; 354pp; German.

XX

CC This invention describes a novel method nucleic acid (I) and its encoded

CC polypeptide (II) containing at least one region that encodes an enzymatic

CC activity involved in biosynthesis of spinosyns. (I) are used (i) to

CC identify, inactive or modulate genes involved in the biosynthesis of

CC (II); (ii) to generate a library of polyketide synthases; (iii) for

CC adding forosamine or trimethylrhamsone to a spinosyn or polyketide

CC aglycone; and (iv) for recombinant production of the corresponding

CC enzymes, which are used for production of (II), their precursors or

CC derivatives, including production of transgenic plants that express (II)

CC and thus have increased resistance to insects. (I) are also useful as

CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are

CC macrolides with insecticidal, but not antibacterial, activity, and can

CC also be used to raise specific antibodies, useful for identifying

CC expression clones in a gene bank. Cells transformed with (I) may produce

CC (II) at significantly increased levels or produce new derivatives of

CC (II). This sequence represents an S. spinosa polyketide synthase.

XX

SQ Sequence 2152 AA;

Query Match 31.5%; Score 2273; DB 22; Length 2152;

Best Local Similarity 32.7%; Pred. No. 5.3e-172;

Matches .600; Conservative 199; Mismatches 542; Indels 496; Gaps 32;

QY 11 EDP1AIVGASCRPLPGVVIDLSGFWTLLSGSRDVTGVAPAR-WDAAMAFDPDDADGKTP 69

DB 36 DDP1AIVMSCRVPEQVSSPEDIMQLAAGVDALISEVGDGMDLGVFVDSDBRGTSY 95

QY 70 VTRASFISDVACFDASFFGISPREALRMDPAHRLILEVCWDALENAIAPALVGTETGV 129

DB 96 ACAGGFLQGVSEBFDGFFGISPREALAMDPOORLLLEVAWEFERAGLEORSTRSGV 155

QY 130 FIGIGPSEYEALPQATASAEIDAHGIGTNPVAGISYALGRCGVAVDTAYSSSL 189

DB 156 FVGINGQYASWL--RTPPPAVAGHVLTGGAATAVLSGVAISFEGEPAVVDTCSSSL 213

QY 150 VAVHLACOSLRSGECSTALAGVSLMSLSPSTLWLSKTRALARDRCFAFSNADGFGRG 249

DB 214 VALHLAQAALRAGECDLALAGVVTMSTPKVFLERSRGLAPDRCKSPAAAGDTGWMG 273

QY 250 EGCNAVVLKRLSGARADGDRILAVIRGSAINHDSAGSLTVPNSSQETVLKRALADGC 309

DB 274 EGAGLLLERLSDARRNHEVLAIVRGSAVNOGDASNGLTAPNGSSQORVITQALASAGL 333

QY 310 AASSVGYEAHGTGTTLDGPIEIQALNAVYGLGRDVATPPLIGSVKTMGHREYASGITG 369

DB 334 SVSDVAVEAHGTGRLDPIEAQALITVYGRDRDGRPLWGSVKSNIQHTQALAGVAG 393

QY 370 ILKVLISLQHGQIPAHLAQALNPRISSGDLRLVTRARTPMPDWTFRRAVSSFFGMSG 429

DB 394 VIKMMAMRHQLPRTLIVSESPSEVDSAGTVOLTEMTWPRGCRVRARVSSFGISG 453

QY 430 TNAHVLEBAPPAATCTPPAP-----ERPABLVLISARTASALDAQARLDHLETTYSQ 483

DB 454 TNAHVILEQPPGVPEQSAGPSGSVVDVVPVPMWSGKTPEALSQAALMTVYDERDV 513

QY 484 CLGDVAFSLATTRSMEHRLVAATSREGRLBAALDAQSGTSPGAVSINDSRGKLAF 543

DB 514 SLDVGSIALTRSLADRAVVLGSDRETLGCVALSAGHBSGLTVGSAG-GRIGF 572

QY 544 IFTGGAGQTLGWGRGLYDWSAFREAFD-LCVRL---FNOELDRLEVMMAEAPSVDA 599

DB 573 VFSGGGQGLMGRLYRAFPVFAAFDEACELDAHIGQEIQ--VREVV-----SGDAQ 626

QY 600 LLDQTAFTQPALFTFEYVALAALMSWGVPELVAGHSIGELVACVAGVFSLEDAVFLVA 659

DB 627 LLDRTLMASQGLFALQVGLIKLDSWGVPSVVLGHSVGEELAAFPAAVGSISGARLVA 686

QY 660 ARGRLMOALPPGGAANVSIPEARADVAAVAAPHAASVSIATAVNAPOQVVIAGAGPVHIA 719

DB 687 GARLMQALPSSGGMLAVPAGEBELMSLADQGRVGTAAVNAAGSVVLSGDRVLDILA 746

QY 720 AAMARGARTKALHYSHAHSPLNAPMLAEGRVAESVSYRRPSIVLSNLSGKA-CTDE 778

DB 747 GLLDQGISRLRLRSHAFHSYRMDPMLAEFAELARTVDYRCEVPIVSTLTGLDDDAGR 806

QY 779 VSSPGYWRHAREVVRPFDGVKALHAAGAGTFVEYGPSTLTGLVPACMPDA---RAL 834

DB 807 MSGPYWVROVREPFRFADGVQALVEHDVAIVVEGPGGALSALIQEVAASDHAGRLSA 866

QY 835 LASSRAGDEPARTVLEALGILMAGVLVSMGLFP-SCGRBRLPTPYMOBERVY----- 888

DB 867 VPAMRRNDQAKVNTALAHVNRGCAVDMSSFFAGTAKQIEPTVAFORORVLPVSD 926

QY 889 ----- 888

DB 927 SGDVYAGLAGAHEHLLGAVVPVAGDEVLLTGRISVTRHPMLAEHVLGEYIVAGTALL 986

QY 889 -----IDTKADDAAR 898

DB 987 EIALHAGERLGCERVEELTEBAPVLPERGAIQVQLRVGABENSGRRPMALYSREGA 1046

QY 899 GD-----RRAPGAGH----- 908

DB 1047 HDWTHAATGRALPGGEAAGADLADMPAGALPVDLDEFYRLAELGLEYPFQSLKAAM 1106

QY 909 ---DEV-----EEGG----- 915

```

Db 1107 RQGEVVAEALPGTEDSGFVHPALLDAALHATAVRMDARLPQWEGVSLHAKAPA 1166
Qy 916 -----AVRGDRSA-----RDLHP-- 930
Db 1167 LRVRVPAGDDAKSLVCDGTGRPVISVRLVLRSAARRTGARQAHOARLYRSLWPTV 1226
Qy 931 -----PPESGRRRKV 940
Db 1227 QLPAPSPPSCVLGTSEVSADIQVYPLRLSLTALLDAGAPPGVVIAPTPGGRTADV 1286
Qy 941 E----- 941
Db 1287 RETTRHALDVQWMLSDORLNSRLLYTQGAVAVEPGEPTDLQAALWGLNSTQTEH 1346
Qy 942 -----AAGD-----RPF 949
Db 1347 PDRFVLVDVPEPQLLPALPGVLACGEPQALRRGAHAPRLAGISDDVLFVPDGTGR 1406
Qy 950 LEIDEPGVLDHLVLRTERAP--GLGEVEIADVAGLSFNDVOLALGMVPDDLPGKN 1006
Db 1407 LEATRPGSLDGLAL-VDEPTATAPLGDEVRIMABAAGVNFRLIALGMVPG----- 1458
Qy 1007 PRLILGECAGRIYAVGEGVNLVVGQPVIALSAGAFTHVTTSALVLPFQALSAIEA 1066
Db 1459 -VASISSEGAGVVEITGGVTGLAPDRVMGMIIPKAFGLAVADHRVYTRIIPAGWSPARA 1517
Qy 1067 AAMPVAVLTAWYALDRIRARLQGERVLIHAATGVGLAVALQWAGVGAEVHATAGTPEKR 1126
Db 1518 ASVPVFLVLTAYALVDLGLRGEBSILVHSAAGVGMAIQARHIGAEVATATASEDKWQ 1577
Qy 1127 AYLESIGVRYVSDSRDFVADVRAMTGGEGVDVVLNLSIGELIDKSFNLLRSHGRPVYL 1186
Db 1578 AV--ELSEHHLASRTCEBQOFLGATGGRGVDVVLNLSAGFPALASLRMLPRGGRFLEL 1635
Qy 1137 GKRDYADNOLG-LRPFRLNLSFSLVDLRGMMLERPARVRLLELGLIAAGVTPPTPT 1245
Db 1636 GKTVDREDVEVADAHPGVSYQAFDTVE-----AGPORGEMHLELVLEFEGRVLEPLPV 1689
Qy 1246 ATLPIARVADAFRMAOQHGLKVLTLTGDPREVQIRIPTHAGAP----STGDRDLDR 1300
Db 1650 TAMDVROAPEALRHLSQARHVGKLVLTMP-----PWMDAGVLTGTG----- 1734
Qy 1301 LASAAPAAALAEAFRTQVSOVLRTPEIKVGAE--ALFTLIGMD-SIMAVELNRRTIEA 1357
Db 1735 -ALGAEVARHLVIERGVNLVIVSRGRPAASGAELVQLTIVGAEVSLOACDVADR-ET 1792
Qy 1358 SUKTKLSTTFSTSPNIALLAQNLIDALATALSLEEV 1394
Db 1793 LAKVLASIPDEHPLTAVVHAAGVLDVGVESSILVERL 1829

RESULT 12
AAE10142
ID AAE10142 standard; Protein; 7068 AA.
XX AAE10142;
AC
XX
XX 29-NOV-2001 (first entry)
XX
XX Streptomyces noursei nystatin gene, NysI (partial).
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; PKS type I.
XX
XX Streptomyces noursei.
OS
XX
XX Key location/Qualifiers
XX FT 34..448
XX FT /label= KS9 domain
XX FT /note= "ketosynthase (KS) domain"
XX FT 572..890
XX FT /label= AT9 domain
XX FT /note= "Acyltransferase (AT) domain"

```

```

FT Domain 904..1123
FT /label= DH9 domain
FT /note= "Dehydratase (DH) domain"
FT 1443..1686
FT Domain
FT /label= KR9 domain
FT /note= "Ketoreductase (KR) domain"
FT 1720..1793
FT Domain
FT /label= ACP9 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 1813..2236
FT Domain
FT /label= KS10 domain
FT /note= "ketosynthase (KS) domain"
FT 2346..2664
FT Domain
FT /label= AT10 domain
FT /note= "Acyltransferase (AT) domain"
FT 2678..2890
FT Domain
FT /label= DH10 domain
FT /note= "Dehydratase (DH) domain"
FT 2983..3229
FT Domain
FT /label= KR10 domain
FT /note= "Ketoreductase (KR) domain"
FT 3265..3339
FT Domain
FT /label= ACP10 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 3358..3780
FT Domain
FT /label= KS11 domain
FT /note= "ketosynthase (KS) domain"
FT 3898..4217
FT Domain
FT /label= AT11 domain
FT /note= "Acyltransferase (AT) domain"
FT 4231..4432
FT Domain
FT /label= DH11 domain
FT /note= "Dehydratase (DH) domain"
FT 4523..4770
FT Domain
FT /label= KR11 domain
FT /note= "Ketoreductase (KR) domain"
FT 4806..4879
FT Domain
FT /label= ACP11 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 4801..5325
FT Domain
FT /label= KS12 domain
FT /note= "ketosynthase (KS) domain"
FT 5432..5754
FT Domain
FT /label= AT12 domain
FT /note= "Acyltransferase (AT) domain"
FT 5768..5977
FT Domain
FT /label= DH12 domain
FT /note= "Dehydratase (DH) domain"
FT 6068..6315
FT Domain
FT /label= KR12 domain
FT /note= "Ketoreductase (KR) domain"
FT 6348..6421
FT Domain
FT /label= ACP12 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 6454..6873
FT Domain
FT /label= KS13 domain
FT /note= "ketosynthase (KS) domain"

WO200159126-A2.
16-AUG-2001.
08-FEB-2001; 2001WO-GB00509.
08-FEB-2000; 2000GB-0002840.
10-APR-2000; 2000GB-0008786.
14-APR-2000; 2000GB-0009387.
XX
XX (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNF-) SINTEF STITTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.

```

PA (ZOTC/) ZOTCHEV S. B.  
 PA (SEKU/) SEKUROVA O. N.  
 PA (FJAE/) FJAEVRIK E.  
 PA (BRAU/) BRAUTASET T.  
 PA (STRO/) STROM A. R.  
 PI Zolchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
 PI Valla S, Ellingsegen TE, Sletta H, Gulliksen O;  
 XX WPI, 2001-557614/62.  
 DR N-PSDB; AAD17185.  
 XX  
 XX New nystatin polyketide synthase polynucleotides and polypeptides,  
 PT useful as antibiotics and antifungals -  
 PS Claim 15, Page 181-185; 266pp; English.  
 XX  
 XX The present invention relates to the cloning and sequencing of the gene  
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme  
 CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.  
 CC The nystatin PKS is useful as antifungal antibiotics. The present  
 CC sequence is a PKS type I encoding Streptomyces noursei  
 CC nystatin gene, NysI (partial).  
 XX  
 XX Sequence 7068 AA;  
 SQ  
 Query Match 31.4%; Score 2260.5; DB 22; Length 7068;  
 Best Local Similarity 38.0%; Pred. No. 3.5e-170;  
 Matches 614; Conservative 170; Mismatches 544; Indels 287; Gaps 39;  
 QY 10 AEDPIVIVGASGRLPGGVIDLSEFWTLLESGRPTVGRVPAER-WDAAWPDPDPDAPGKT 68  
 DB 4899 AADPIVIVGASGRLPGGVIDLSEFWTLLESGRPTVGRVPAER-WDAAWPDPDPDAPGKT 4958  
 QY 69 PYTRASFLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENAATAIPASLVGTETG 128  
 DB 4959 YSVQGGFLRDVAEFDPGFGISPREALRMDPQRLLETFAMEAFHAGIDPVGQGRSRTG 5018  
 QY 129 VFIIGTPESEYEAALPDATSAEIDAGGIGTGPVAGAGRI SYALGIRGCVAVDTAYSSG 188  
 DB 5019 TFGASVODYASGVSPNSEGSG---EGHMITGTLSSVLSGVSYLFGEGGAVTLDTACSSS 5075  
 QY 189 LVAVHLACOSLRGECSTLALAGVSLMSPSTLWMLSKTRALARDRCKAFSAEADGFR 248  
 DB 5076 LVAHMLACOSLRNGESSLALAGVSLMSPSTLWMLSKTRALARDRCKAFSAEADGFR 5135  
 QY 249 GEGCAVAVLKRSLGABADGRI LAVIRGSAINHAGSSGGLTVPNGSSQSEITLKRALADAG 308  
 DB 5136 AEGVGLVLERLSADANGHOVLAVIRGSAVNDGASNGLTAPNGPSQQRVIRKQALANSA 5195  
 QY 309 CAASVGVYBAHGTGTTLDPTFIQALNAVYGLGRVATPPLIGSVKTNLGHREVASGIT 368  
 DB 5196 VAPGDIDVIEGHGTALDPTIEAQLLTYGODRAPERPLLIGSVKTNLGHREVASGIT 5255  
 QY 369 GLLKVVLSLOHGOIPALHQAOLNPRIWMDLTLTVRATPMPDNTPRBAGVSSFGMS 428  
 DB 5256 SVIKLVRAIOBGVVPVSLHIDRPSTHYDSSGAIIGLITKTPPERGPRRAAVSSFGIS 5315  
 QY 429 GTNAHVLEAPAAATCTPPAPERPAELV---LSAFTASALDAQAARLDHLETTYSQCL 485  
 DB 5316 GTNVHTILIEQAPDAEPAPRA-DEPRDGLVPLLSSGGEALRAQAARLLAFVEERDEAHL 5374  
 QY 486 GIVAFSLATTRSMERLVAATSRGRLAALDAAAGCTSPAVASIASDSSGKLAF 545  
 DB 5375 TDLAHSLATSRALERRAAVIAADRITLGRGLALDGDGRDPGLVGGTA--GGRGRTAF 5432  
 QY 546 TGGGAQTLGKRGGLVYVWSAFREAFDLCTRLFNQGLDRPLREYMALEPASVDALLDOTA 605  
 DB 5433 TGGGSORPGKRGRLRYVFDALDEVLARLDGDGRPLREVLFAAPDSAEALDLDRIG 5492  
 QY 606 FTOPALFTEYALAAALMRSGVPELVAGHSIGELVAAACVAGVSLDPAVFLVAARGRL 665  
 DB 5493 YAGPALFAVAVVALLFRLLTSGGLTPOVLAHSGVGLAAAVAGVLSDDACTVVAARGRL 5552

QY 666 QALPAGAMVSTEARPEADVAAAAPHASVSTIAANPADOVVIAGAGCPVHAIAAAMAR 725  
 DB 5553 QALPAGAMVSTEARPEADVAAAAPHASVSTIAANPADOVVIAGAGCPVHAIAAAMAR 5612  
 QY 726 GARTKALVSHAFSHPSPLAAMLEAFGRVAESYSRPSIVLSNLSGRAC-T-DEVSSEPGY 784  
 DB 5613 GARTKALVSHAFSHPSPLAAMLEAFGRVAESYSRPSIVLSNLSGRAC-T-DEVSSEPGY 5672  
 QY 785 WYRHAREVVRADGVKALHAAG-AGTVEVGKSTLLGLVPACMPD-ARPALLASSRAGR 842  
 DB 5673 WYRHAREVVRADGVKALHAAG-AGTVEVGKSTLLGLVPACMPD-ARPALLASSRAGR 5732  
 QY 843 DEPAVLEALGMLAVGVLSMAGL-FPS-GGRRVPLPYPMORERYWIDTA- 893  
 DB 5733 DEPAVLEALGMLAVGVLSMAGL-FPS-GGRRVPLPYPMORERYWIDTA- 5792  
 QY 894 -----DDAA----- 897  
 DB 5793 SALDAEFMAVERDDVAAALASLDDATVTYAMVPALAMRRRREGOTELDSWRYVTWK 5852  
 QY 898 -RGDRRPA- -----GHDEVEGAVRGDRRA- -----RLDHPPEEG--R 936  
 DB 5853 PRGCATAPAAULTGRWLVLVPHDQRODAPTAWADVETALGTTVRLVTTTDRALA 5912  
 QY 937 REKVAAGDR--PERLEIDEPGLDHLVRYTERRARG- -----LGEVEIVDAAGSFNDVQ 991  
 DB 5913 ARITBAAGDQGF- -----SGVLSLPLATGAGHPGAPALTLTTTAVQALGDGIDAP 5966  
 QY 992 L-----ALG-----WVPDDL- -----GKPNPULLGEGCAGRIYA 1021  
 DB 5967 LKNYTRGAVAVGRAGVATAPBEGAAMVGLGRAVLELPAFGSTLDPATLPGQAARLRA 6026  
 QY 1022 VEGGVNGLVGGPVIALSAGAFHTVTSALVLEPRPALSAIEAAMPVAVLTAVYALD 1081  
 DB 6027 VLAARD-----GEDAVAL-----RPSGVFLRLAHAPAGDTRTAFD 6064  
 QY 1082 RIARLOPGRVLIHAATGGVGLAAVQMAHGAEVH-----ATAGTEPKARYLESL 1132  
 DB 6065 PAA-----GTVLITGCTGIGIIGHVARLARDGA-THLLTSRRCGAPAGADALRALEBEL 6118  
 QY 1133 GVR-----ALG-----YVSDRS--DRFVADVAM 1152  
 DB 6119 GARVITLACDADRALALALAEPLDPDAPLCAVFTAGVEDHVVDALTPENFAVLRAK 6178  
 QY 1153 TCGEGVDVYVLSLSEELIDKSNLRSRGRFVELKRCYADNOLGLAPPLRN-LSFSLV 1211  
 DB 6179 TVAAH-----HLHETADLDLAAVFLFSSTAGVYGAAGQGNVYAAANHALALAEHRRSHGIL 6235  
 QY 1212 DL-----RGWLERPARVRLLEELGLIAGVFTPPITATLPIARVADAPRMAQ 1262  
 DB 6236 ALSVAMGPMAGSMVAD-----AAELTDVRRGCF--EPLAEP-----AVRALLR 6279  
 QY 1263 ACHLKLVLTLGD-----PEVOIRIPTHAGAGSTGDRDLDR 1300  
 DB 6280 ALENDOTVYALADIMERFORAPFAVRPLPVADLPETGRATPATA-TGAATG--LRQ 6335  
 QY 1301 LASAAPAAALAEFLRTQVSQV-----RTPELKVGAELFTLGMDSLMAVETLRRI 1355  
 DB 6336 LAELPEHERPAVVDLRTQVAAVIGHADPRVE-----DDHAFRDLGDFDSLITLELRNAL 6391  
 QY 1356 EASLKLKSTFSLTSPNIALLAQNLDALA-----TALSLEVAENLELRAGVQ 1404  
 DB 6392 NATGSLSPATLVYDLPTRERADTLLEBLGTLPTDTAAVYASTASRKLSSAFE 6446  
 RESULT 13  
 AAE10144  
 ID AAE10144 standard; Protein; 9477 AA.  
 XX  
 XX AAE10144;  
 AC  
 XX  
 DT 29-NOV-2001 (first entry)

XX Streptomyces noursei nystatin gene, NysI (complete).  
DE Polypeptide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
KM antifungal; antibiotic; PKS type I.  
XX Streptomyces noursei.  
OS  
FH Key  
FT Location/Qualifiers  
FT 41..464  
FT /label= KS15 domain  
FT /note= "Ketosynthase (KS) domain"  
FT 578..889  
FT /label= AT15 domain  
FT /note= "Acyltransferase (AT) domain"  
FT 903..1102  
FT /label= DH15 domain  
FT /note= "Dehydratase (DH) domain"  
FT 1446..1731  
FT /label= ER15 domain  
FT /note= "Enoylreductase (ER) domain"  
FT 1740..1988  
FT /label= KR15 domain  
FT /note= "Ketoreductase (KR) domain"  
FT 2023..2096  
FT /label= ACP15 domain  
FT /note= "Acyl carrier protein (ACP) domain"  
FT 2117..2538  
FT /label= KS16 domain  
FT /note= "Ketosynthase (KS) domain"  
FT 2635..2953  
FT /label= AT16 domain  
FT /note= "Acyltransferase (AT) domain"  
FT 2967..3167  
FT /label= DH16 domain  
FT /note= "Dehydratase (DH) domain"  
FT 3257..3500  
FT /label= KR16 domain  
FT /note= "Ketoreductase (KR) domain"  
FT 3539..3612  
FT /label= ACP16 domain  
FT /note= "Acyl carrier protein (ACP) domain"  
FT 3634..4057  
FT /label= KS17 domain  
FT /note= "Ketosynthase (KS) domain"  
FT 4153..4472  
FT /label= AT17 domain  
FT /note= "Acyltransferase (AT) domain"  
FT 4486..4725  
FT /label= DH17 domain  
FT /note= "Dehydratase (DH) domain"  
FT 4997..5245  
FT /label= KR17 domain  
FT /note= "Ketoreductase (KR) domain"  
FT 5277..5350  
FT /label= ACP17 domain  
FT /note= "Acyl carrier protein (ACP) domain"  
XX  
PN W0200159126-A2.  
PD 16-AUG-2001.  
XX  
PF 08-FEB-2001; 2001WO-GB00509.  
XX  
PR 08-FEB-2000; 2000GB-0002840.  
PR 10-APR-2000; 2000GB-0008786.  
PR 14-APR-2000; 2000GB-0009387.  
XX  
PA (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPLIGE.  
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.  
PA (ALPH-) ALPHARMA AS.  
PA (SINV-) SINVENT AS.  
PA (DZIE/) DZIELEMSKA H.

PA (ZOTC/) ZOTCHEV S B.  
PA (SEKU/) SEKUROVA O N.  
PA (FJAE/) FJAEVRIK E.  
PA (BRAU/) BRAUTASET T.  
PA (STRO/) STROM A R.  
XX  
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;  
PI Walla S, Ellingsen TE, Sletta H, Gulliksen O;  
XX WPI: 2001-557614/62.  
DR N-PSDB: AAD17186.  
XX  
PT New nystatin polypeptide synthase polynucleotides and polypeptides,  
PT useful as antibiotics and antifungals -  
XX  
PS Claim 15; Page 255-260; 266pp; English.  
XX  
CC The present invention relates to the cloning and sequencing of the gene  
CC cluster encoding a modular type I polypeptide synthase (PKS) enzyme  
CC involved in the biosynthesis of the macrolide antibiotic nystatin.  
CC The nystatin PKS is useful as antifungal antibiotics. The present  
CC sequence is a PKS type I encoding Streptomyces noursei  
CC nystatin gene, NysI (complete).  
XX  
SQ Sequence 9477 AA;  
Query Match 31.4%; Score 2260.5; DB 22; Length 9477;  
Best Local Similarity 38.0%; Pred. No. 5.5e-170; Indels 287; Gaps 39;  
Matches 614; Conservative 170; Mismatches 544;  
QY 10 AEDPIAVGASCLPGGVIDLSGFWTLLEGSRTVGRVPAER-WDAAMFDPDPADQKT 68  
DB 4899 ADPIAVGMSCHRYPGGVGSAEDLMRIALDEVDALISGFADRDGADAGLVDPDRDQRT 4958  
QY 69 PVTBASFLSVACPDASFPFISPREALRMPARLLLEVCEALENNAIPSLVETETG 128  
DB 4959 YSVQGFRLRDVAFFDPFGFISPREALSMPOORLLLETWEAFEHAGIDPVGGRGSRGT 5018  
QY 129 VFIGIGPSEYEALPONTASAEIDAHGGLTWPVGGRISYVALGLGPCVAVDTAYSSS 186  
DB 5019 TFWGASTQDYASGVPSSEGS---EGHMITTLLSSVLSGRVSYLFGFEGPAVTLDTACSSS 5075  
QY 189 LVAVHLACOSLRSGECSTALAGVSIMLSPSTLVMSKTRALARDGRCKAPSAEADGFCR 248  
DB 5076 LVAMHLACOSLRNGESSLALAGVSIINSTWSPFGSRQALAEDEGRCKAYADAGDMTL 5135  
QY 249 GEGCAVVVTKRLSGARADGRRITAVTRGSAINHDASSGGLTVNGSSQGEIVLKRALADAG 308  
DB 5136 AEGVGLVLERLSDPARANGHQLAVIRGSAVNDGASNGLTAPNGPQORVIRQALANSA 5195  
QY 309 CAASVGVVEAHGVTTLGDPPIEIOALNAVYGGRDVATPLLGSVKTNGHPYASGIT 368  
DB 5196 VAPGDIVLEGHGTGLAGDPIEIOALNAVYGGRDVATPLLGSVKTNGHPYASGIT 5255  
QY 369 GLKVVVLSQHQOIPALHQAOLNPRISWGLDLRLTVTRARTPMDWTPPRRAGVSSFGMS 428  
DB 5256 SVIKLVRALQEGVVPVPSLHIDRPSITHVDMSSGAILGILTERTPMPETPRPRAAASSGIS 5315  
QY 429 GTNAHVLEBAPAAATCTPPAPERPABLIV--LSARTASALDAQAARLRLDHLLETYSQCL 485  
DB 5316 GTNVHTLIEQAPADEAPTTPA-DEPRDGLVPLLSGREBALRAQNALFLFVEERRPRAHL 5374  
QY 486 GDVAFSLATRSAMEHRLAAVATSRBETRAALAAAGQGRSPGAVRSIAASSRGKLAFLF 545  
DB 5375 TDLAHSLATSRALDERAAVIAADRDVTLTGLRLSLSGRDPGLVQGTQ--GKRITAPLF 5432  
QY 546 TGGCAQTLAGRGGLYDVWSAFREAFDLCVRLFNQELDRPLREYVMAEPASVDAALDOTA 605  
DB 5433 TGGGSGRPQGRGLHRYPIVFADALDEVLARLDGDPRIREVLFAPADABABALDRTG 5492  
QY 606 FTQPALFTFEYALALMRSGVEPELVAGHSIGELVAAVCYAGVPSLEDAVFLVAARGRLM 665  
DB 5493 YQAPALFAVEVALFRLLTSGWLTPTDYLAGHSVGLAAAHVAGVLSDDACTLVAAARGRLM 5552

```

QY 666 QALPGAMVSIIEAPADVAANAAPHASVSIANVAPDOVVIAGAGPYHAIAMAMAR 725
DB 5553 QALPGAMVSIIEAPADVAANAAPHASVSIANVAPDOVVIAGAGPYHAIAMAMAR 725
QY 726 GARTYALVSHAFHSPMLAEMLEAFGRVAESVSYRPSIVLSNLSGKACT-DEVSSPGY 784
DB 5613 GRTYALVSHAFHSPMLAEMLEAFGRVAESVSYRPSIVLSNLSGKACT-DEVSSPGY 784
QY 785 WVRHAREVVRPADGVKALHAAG-AGTFVEVGPSTLGLVPAQMPD-ARPAIILASSRAGR 842
DB 5673 WVGHVRAVVRPADGVKALHAAG-AGTFVEVGPSTLGLVPAQMPD-ARPAIILASSRAGR 842
QY 843 DEPARVLEALGLMVAVGLVSWAGLPPS-GGRVRLVPTPYMWRREYIMDTKA----- 893
DB 5733 DEPARVLEALGLMVAVGLVSWAGLPPS-GGRVRLVPTPYMWRREYIMDTKA----- 893
QY 894 -----DDAA----- 897
DB 5793 SALDEFMAAVERDVAALASLDDATVTAMVPAITAMRRRREGOTELDSWRVYTWK 5852
QY 898 -RGDRRAPGA-----GHDEVEGGAVRGDRRSA-----RLDHPPESSG--R 936
DB 5853 PRGATAPALALGRWLIVPHDQRODDATAMADVETALGTTVRLVTTTDRALA 5912
QY 937 REKVAADGR-PFRLEIDEPGLDHLVRYTERRAPG-----LGEVIAVDAAGLSFNDVQ 991
DB 5913 REKVAADGR-PFRLEIDEPGLDHLVRYTERRAPG-----LGEVIAVDAAGLSFNDVQ 991
QY 992 L-----ALG-----WVDDLP-----GKPNPPLLGGECAGRIYA 1021
DB 5967 LWNVTRGAVAVRAGQVTAPEQAAVWGLGRAVLELPAFGTLDLPALDQAAARLRA 6026
QY 1022 VGEVNGLVGQPVIALSAGAFATHTTSALVLRPQALSAIEAAMPVAVLTAMVALD 1081
DB 6027 VLAATD-----GEDAVAL-----RPSGVFLRLHAHAPGDTARTAFD 6064
QY 1082 RIARQPERVLIHAATGCVGLAIAVQMAOHVGAEN-----ATAGTPKRAYLESL 1132
DB 6065 PAA-----GTVALITGTGTGIGHVARRLARDA-THLLITSRGPAAPADALRLLEL 6118
QY 1133 GVR-----YVDSRS-DRFVADVRAM 1152
DB 6119 GARVTLAACAADADRALAALLAELPDDAPLCVHTTAGVVEHVVDALTPENFAVLRK 6178
QY 1153 TGGEGVDVNLNLSGELIDKSFNLLRSHGRFVELKRDVADNOLGLRPFLRN-LSFSLV 1211
DB 6179 TVAAH-----HHELTADLDLAELFSSTAGVLAAGQGVYAAANHLDLAEHRRSHGLT 6235
QY 1212 DL-----RGMLEERPARVRLLELGLIAGVFTPPPIATLPIARVADARSMQ 1262
DB 6236 ALSVAMGFWAGSGMAD-----AAELTDRVRGGF--EPLAEP-----AVRALLR 6279
QY 1263 AQHLGKLVLTGD-----PEVOIRIPTHAGAGPSTGDDLLDR 1300
DB 6280 AIENDDTVALADIMWERFORAFVAVRPLPFVADDEPGRATPARA-TGAAGC---LACQ 6335
QY 1301 LASAAPARPAALAEALFRTQVSQV-----RTPEIKVGAELFTLGMDSLMAVELRNI 1355
DB 6336 LAELPEHRRPAVADLRLTQVAAVLGHADPRIVE-----DDHAFRLDGFESTLLELRNAL 6391
QY 1356 EASLTKLSTFELSTSPNIALLAQULDLA-----TALSIERVAENLEPRGVQ 1404
DB 6392 NNAATGLSLPATLVYDLPTPREWADFLAELLGTLPTDTAATVASTASPRLSHSPF 6446

```

RESULT 14  
 AAU10701  
 ID AAU10701 standard; Protein; 6095 AA.  
 AC AAU10701;  
 XX  
 DT 12-MAR-2002 (first entry)

```

XX DE Amino acid sequence encoded by S. cellulosum PKS cosmid pKOS28-26, ORF2.
XX KW Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
XX KM acyl transferase domain; dehydratase domain; ketoreductase domain;
XX KW acyl carrier protein domain; pKOS28-26.
XX OS Sorangium cellulosum.
XX FH Location/Qualifiers
XX FT 34..457
XX FT /label= Ketosynthase_domain
XX FT 568..895
XX FT /label= Acyltransferase_domain
XX FT 922..1100
XX FT /label= Dehydratase_domain
XX FT 1429..1614
XX FT /label= Ketoreductase_domain
XX FT 1706..1789
XX FT /label= Acyl_carrier_protein_domain
XX FT 1812..2235
XX FT /label= Ketosynthase_domain
XX FT 2341..2673
XX FT /label= Acyltransferase_domain
XX FT 2978..3162
XX FT /label= Ketoreductase_domain
XX FT 3250..3333
XX FT /label= Acyl_carrier_protein_domain
XX FT 3356..3781
XX FT /label= Ketosynthase_domain
XX FT 3887..4219
XX FT /label= Acyltransferase_domain
XX FT 4246..4424
XX FT /label= Dehydratase_domain
XX FT 4754..4939
XX FT /label= Ketoreductase_domain
XX FT 5031..5114
XX FT /label= Acyl_carrier_protein_domain
XX FT 5134..5558
XX FT /label= Ketosynthase_domain
XX FT 5664..5991
XX FT /label= Acyltransferase_domain
XX FT 6018..6095
XX FT /label= Dehydratase_domain
XX PN US6280999-B1.
XX PD 28-AUG-2001.
XX PF 31-AUG-1998; 98US-0144085.
XX PR 22-JAN-1998; 98US-0010809.
XX PA (KOSA-) KOSAN BIOSCIENCE.
XX PI Gustafson C, Betlach MC, Ashley G, Julien B, Ziemann R;
XX WP1; 2001-606536/69.
XX DR Novel purified, isolated DNA molecule from Sorangium cellulosum having
XX PT polyketide open reading frame encoding modules with one or more domains
XX PT such as ketosynthase, acyl transferase and acyl carrier protein domains
XX PT
XX PS Claim 1; Fig 4; 72pp; English.
XX CC The present invention relates to the isolation of novel Sorangium
XX CC cellulosum polyketide synthases (PKS), and the polynucleotide sequences
XX CC encoding them. The polyketide synthases include catalytic domains such
XX CC as ketosynthase domain, acyl transferase domain, dehydratase domain,
XX CC ketoreductase domain and acyl carrier protein domain. A host cell
XX CC comprising a PKS ORF (open reading frame) which encodes one or more
XX CC more PKS domains is useful for producing polyketide synthases from which

```



polyketides can be produced. The host cells are useful for constructing a library, where each individual colony of the library represents a colony with the ability to produce a particular PKS synthase and ultimately a particular polyketide. The polyketides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity. Colonies in the library are also induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of candidate polyketides which can be screened for binding to desired targets such as receptors, signalling proteins, etc. The present sequence represents the amino acid sequence of encoded by ORF2 of S. cellulosum PKS cosmid pKOS28-26. Note: The present sequence is said to correspond to S. cellulosum PKS functional domains or domain subsets encoded by ORF2 of the cosmid pKOS28-26 (AA517367).

CC Sequence 6095 AA;

Query Match 31.2%; Score 2250; DB 22; Length 6095;

Best Local Similarity 35.0%; Pred. No. 1.9e-165; Matches 630; Conservative 205; Mismatches 550; Indels 414; Gaps 50;

```
7 ERAADPIATVAGACRLPGVLDLGGFWTLTSGSDTGRVPAERMDAAAFDDPPAPG 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3351 EEGSTPIATVGVGRNPGGASDLSEFWQVLEERDILRPIDQDFDEALYDPPDKG 3410
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 KTPVTRASFISDVAACPDAPFGISPREALRMDPAHRLLEVCWEALENAAIPASLVGTE 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3411 KTYRNASLDDVASFDPGFGISPREABPMDPQRLILETAMSLBEDAGVRPEHLKSD 3470
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 TGVFTIGSPSEYEALPOATSAEIDAHGIGLTPSVAGRISTALGIRGCVAVDTRYAS 186
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3471 TGVFVGAAPSEYAS---YRGSANEDAYALGTALSFAGGVAAHGLGQGAIVSDTACS 3527
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 SSVAAVHLACSLSGECSTALAGVSLMLSPSTLWMLSKTRALARDGRCAFSADNGF 246
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3328 SSVAAVHLACALNRGCEVALLAGVQLANPAGVILSRTRALSPDRCALFQADNGY 3387
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
247 GRGEGCAVVVLKRLSGARADDRILAVIRGSAINHGDASGLTYPNSSQOEIVLKRALAD 306
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3388 GRGEGVGLVLMRLSEAOQOKRVLGVRGTAVNODGASSGITAPNGTAQOKVRAALRN 3647
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
307 AGCAASVGVYEAAGTGTTLDDPIEIOALNAVYGLGRVATPLIGSVKTMHGHEVYASG 366
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3648 AGLEPASIDVVECHGTGTSLDPIEQALGAVYGGGRDMARPLQAGVKSNIHGHESAG 3707
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
367 ITGLKVLVLSHQGIPAHILHAQALNPRISMGDLRLTYTRARTPMP--DNMTPRAGVSS 424
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3708 IAGVCKTLAARVYSLPATLHSSPPNPRIPWENLPEVQVDRLTTPWPRABEPPRRAGVSS 3767
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 FGMGNTAAVVLLEAPATCTPP--APERPAEL--LVLSARTASALDAQAARLRLHLETYP 481
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3768 FGISGTAAVHLEAPAREPAREPVEBAAPLPLVLSGRDEAVNQAQGMWAKMEHSG 3827
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
482 SQCLDVAFSIATTRSAHEHRLAAVATSRBGLRAALDAAQOGTSPGAVRSIADSSRGL 541
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3828 EVGMSDVVTRALHRTHEESPASVLAASAAGVEGLRALSGRDPAAVVSTAKRG--GKL 3886
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
542 AFLFTGGAGOTLGMGRGLYDVMSAFREAFDLCTVLFNQLDRPLREVMMAEPASVDALL 601
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3887 AALFTGGOSQRLGMGRGLYEVYVFRALFDEVCEALDHLDRGLREVFFAAGSEEGQL 3946
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
602 DQATFOTPALFTFEYALAAALWRSWGVPELVAHSGISGELVAACVAGVSELDFAVLVAR 661
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3947 ERTETYPGLFALVALYRWESWGLKRALHSGISGELSSAAHAYAGVSLADAKLYAR 4006
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
662 GRMLQALPAGGAMVSTEAPEADVAANAAPHAAS--VSTAAVNAADQVVIAGAGPVHAIA 719
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4007 GRMLQGEAGGAMVSEASBEVORALSEVAGOGELSTAGLANAPQVTLSDDEAAVLAVA 4066
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
720 AAMARGARTKALVHSHAFHSLPMLAEARGVAESYSYRPSLVYVSNLSGKACTDE 778
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4067 RLRLAQGRTRRLKLVSHAFFAHMDGMLBEFGKVARRECTIARPOLAVVSVTGLGGEBA 4126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 779 VSSPGVWVHAREVVRADGVKALHAAGACTFEVEGPKSTLGLVPAQMPD-ARPAALAS 837
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4127 LMSAEYWRQVREAVRFLDQKRTIAAGVSTYVCEGDPVLCALGACGLPGCAEATFVAS 4186
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 838 SRARDEPATTLEALGLMNVAGVSWAGLEFPS--GGRVRPLPTYPMQERERWID----- 890
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4187 LRROBERRALATVATVATVVOGHEVDVAQVLSGRGSRFVELPTVAFQORWLEAPKART 4246
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 891 -----TK-----ADDAAGDRRAPAGNDE 910
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4247 DVGSAGLRESGHPILGAATKLADDDHLFTGRSLSGQPMWRDAVREGVVFPFGMLDL 4306
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 911 -VEEGAVRGGRDRSARLDHP-----PPESGR-----EKVBAADR 946
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4307 ALAAGRTVSGGALSELITSEPLMAEDVAVRLQLSVGAPDAAGRRARFGLYQPEQPGDA 4366
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 947 PF-----RLRIDE-----FQVLDHL-----VLR 964
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4367 PMVGHATGVLTDELATLSGELDELTPVPQAEAVDLSGFYERLHGRGARYGPAFQGLVE 4426
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 965 VTERRAPGLGEVEIAYDA-----AGLSFNDVQLAGVPPD--LP----- 1002
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4427 LSRRDAPFGGRVULPKDATTSAEDYGVHPALMDAALHTTMAAFREVASPDDVLLPFSMSD 4486
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1003 -----GKNPPPL--LLGECAGRIVAVEGVNGVLVQGPVIALSAGAF----- 1043
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4487 VALHATGASELRVRLTAGGRDSAOAAASLRVTD--AAGGVV--SVGALLHRTATAEQL 4542
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1044 -ATHV-----TTSALVLPFRQOLSAIEAAMPVAYLTAWAYLD 1081
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4543 RAAATHABAQHLVYRDFQLVSLVBAKSVYDLSLVIRABEGRGL--GEALGVALIGLDL-- 4600
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1082 RIARLQPG--ERVL-----HAATG-----GVGIA-- 1104
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4601 -LARIEGTLPREVLVDMTNAGSSQSDMYISSEHAIGQALSLQAMLSERLGEVLYW 4659
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1105 ---AVQMAOHVGAE--VHA-----TAGT--DEKRAYLESIGVRVYSD----- 1139
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4660 VTRDAVAAAPDQGVQDIAHAFLMGLVTRTARSEHERRLRILDVGTPELDGILLARALATA 4719
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1140 -----SRDRFADVRAMTGGECVD-----VYVNSLSGELIDK--STNLRSH 1180
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4720 TEPELALRGAMAAARLVRRVPAABEGULTPARGLDPTGTVLVTGTGELGQAVAEHLVRAH 4779
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1181 G--RFEVLEGKRDQVYADNQLGRPEPLRMISFSLVDLRGML--ERPARVRALLB----- 1229
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4780 GVRHLVLTSSRGLEAPGAPGVQALEKLGAEYTVVAAACDVSKREVARVLAGIEAAHPLT 4839
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1230 ---ELGLIAAGVF--TPPEIATLPIARV----- 1253
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4840 AVHLAGVLDGVTIATQTPRELRSVLPAKVNAGALHLHETBEDDLAFLVFPSSMSGTIGT 4899
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1254 -----ADAFPS-----MAQOHLGKLVLT----- 1272
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4900 AGGSNYAANSFLDPAFAHRRSRGLAATSLAMGFMAQTGVMTHLGEALISRIORAGLV 4959
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1273 -----IGDPEVOIRIPTH-----AGAP-----ST 1292
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4960 PIRVEBGLSLDAAALRPENSL--VPAHLDLAQMGKLEASGELPALLRALIRPGLRASS 5018
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1293 GDRD---LIDRLASAAPARAAALAEFLRTQVSOVLRTPEIK--VGAELPTRLGMDSLM 1347
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5019 ATRKEASALMERUSELPEABERLSLVELVAEVAAYVGLPRSEAVAAVDQVLKDGGLDSLM 5078
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1348 AVELKRIEASLKUKSTFTPLSTSPNTALLAQUNLDLATRLSLERYAAENLRAGVOND 1406
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5079 AVELRSRLSARAEIPLPATVLFVDFPTPRAVAEEL---LRQAFSKQOVTAARARRRTEED 5134
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 15
AA58577
ID AA58577 standard; Protein; 3798 AA.
XX
```

AC AAY58577;  
 XX  
 XX 10-APR-2000 (first entry)  
 DT  
 XX Sorangium cellulosum type I polyketide synthase EPOS D.  
 DE  
 XX EPOS D: type I polyketide synthase; epothonone biosynthesis;  
 KW polyketide backbone formation; taxol substitute; anticancer.  
 XX  
 OS Sorangium cellulosum.  
 MO966028-A2.  
 XX  
 XX 23-DEC-1999.  
 PD  
 XX 16-JUN-1999; 99MO-EP04171.  
 PF  
 XX 18-JUN-1998; 98US-0099504.  
 PR 24-SEP-1998; 98US-0101631.  
 PR 05-FEB-1999; 98US-0118906.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 PI Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;  
 DR WPI: 2000-097741/08.  
 DR N-PSDB; AA255887.  
 XX  
 PT New isolated epothonone synthase genes, used for the recombinant  
 XX production of epothonone for use in cancer therapy  
 ES Claim 12; Page 139-150; 174pp; English.  
 XX  
 XX This sequence represents a Sorangium cellulosum type I polyketide  
 CC synthase, EPOS D, which is one of several epothonone biosynthetic  
 CC enzymes encoded by a 68.75 kb contig. Epothonones A and B are  
 CC 16-membered macrocyclic polyketides with an acylcycetene-derived  
 CC starter unit; polyketides being synthesized from two-carbon building  
 CC blocks, the beta-carbon of which always carries a keto group. Each round  
 CC of two-carbon addition is carried out by a complex of enzymes known as  
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.  
 CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of  
 CC the thiazole ring formation of epothonones, and EPOS B, EPOS C, EPOS D  
 CC and EPOS E (AAY58575-Y85878) are involved in polyketide backbone  
 CC formation. EPO F (AAY58579) is an epothonone macrolactone oxidase, and  
 CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be  
 CC involved in transport. Epothonones mimic the biological activity of  
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic  
 CC compositions. Epothonones exhibit a much lower drop in potency against a  
 CC multiply drug-resistant cell line compared with taxol, and are  
 CC considerably less efficiently exported from such cells by the multidrug  
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of  
 CC epothonones as anticancer agents, they are problematical to produce on a  
 CC large scale. Epothonones are too complex for industrial scale chemical  
 CC synthesis, and Sorangium cellulosum is difficult to ferment, producing  
 CC poor yields of epothonones. The nucleic acids of the invention may be  
 CC used for the recombinant production of epothonones in a heterologous host  
 CC that is more amenable to fermentation.  
 XX  
 SQ Sequence 3798 AA;  
 Query Match 31.1%; Score 2245.5; DB 21; Length 3798;  
 Best Local Similarity 36.2%; Pred. No. 2.1e-169;  
 Matches 605; Conservative 191; Mismatches 495; Indels 379; Gaps 43;  
 QY 9 AAEADIAIVGASCRIPGGVYIDLSGFMTLLSEGRDITGVAPARWPAAMAFDDPDAPGKT 68  
 DB 15159 AADDDIAIVGACRPPGGDEGLTYWRHLAEGNVSTVEPADRMRADWDIDDPVPGRT 1578  
 QY 69 PVTASFLSDVACPDASFGISPRBALRMDPAHRLLEVCWEALENAAIAPASALVTGTG 128  
 DB 1579 YVAKGAFLRDVAISLDAAPFAISPREAMSLDPQGRLLLEVSWEAIRAGDDPALRESATG 1638

QY 129 VFIGIPSEYRAALPQATASAEIDAHGGLGTMPSVGAGRIISYALGLRGCVAVDTAAVSS 188  
 DB 1639 VFVGMIGSEHAERQGLDDDDAAL-LYGTGMLLSVAAGRISFPLGHPTMTVTDACSS 1697  
 QY 189 LVAHLACQSLRSECESTALAGVSLMSPSTLVLSKTRALARDGRCKAFSAEADGPGR 248  
 DB 1698 LVALLHACQSLRSECESTALAGVSSVLLSPSPFVAASMRLLSPDGRKTSAAADGPAR 1757  
 QY 249 GEGCAVVVLRKLSGARADGDRILAVIRGSAINHDCASSGLTPVNGSCOEIVUKRALDAG 308  
 DB 1758 AEGCAVVVLRKLRDQDRDPIILAVRSTAIINHDPSSGLTPVNGSPAOQALRLQALDAG 1817  
 QY 309 CAASVGVYEAHAGCTTGTDPIEIOALNAVYGLRDVATPLLIGSVKTNLGHPEVASGIT 368  
 DB 1818 VAPAEVDFVECHGTGTALGDPLEVOALGAVYGRGPRAPRPLMLGAKNHLGLEAAAGLA 1877  
 QY 369 GLTKVVLSTLOHGOIPAHILHAQALNPRLSWGBDLRLVTTRARTPMDMTPRRAGVSPGMS 428  
 DB 1878 GVLKVLALHEHQIPAOPELDELNPHIPMAELPVAVVRAYPWRPGARPRRAGVSAFGLS 1937  
 QY 429 GTNAHVLEEARPAATCTPPAPERPAELIVISARTASALDAQARLRDLHLETPSOCLDV 488  
 DB 1938 GTNAHVLEEARPAVEPAVAAPERAAELFVLSAKSAAALDQAARLRDLHLEHVEIGLGDV 1997  
 QY 489 AFSIATTSAMEHRLAVALATREGLRALDDAAQOQTSFGVRSIAD-SSRGKLAFLFTG 547  
 DB 1998 AFSIATTSAMEHRLAVALASREALRGALSAQAQHTPPGAVRGASGSAFVKVFPVG 2057  
 QY 548 QGAOTLGKRGGLYDWSAFREAFDLCVRLFNOELDRPLREVMMAEPASVADALDQTAFT 607  
 DB 2058 QGSQWVGKGRKLMAEPIFRALAECDRAIEAGKMSLGLSLADEA--ASQGRIDIV 2114  
 QY 608 QPALFTFEYALALIRSWGVPELVAGHSISELVAAVCYGVPSLEDAYFLVAARGELMOA 667  
 DB 2115 QPVLFAFMEVALSALMRSGVPEAVVGSMEVAHAHVAAGALSIEDAVAIICRSRLLR 2174  
 QY 668 LPAQAMVSIAPAEADVAAVAAPHAASVSIANAAPDQVYIAGAGOPHAIATAAAMARGA 727  
 DB 2175 ISGQEMALVELSBEAPALRGHEGRISLVAVSFSSTVLAGEPALISEVALIATLTAAGV 2234  
 QY 728 RTKALHVSAAHSPPLMAEMLEAFGRVAVESYRRPSIVLVSNGKACTDEVSPPGVWR 787  
 DB 2235 FWRQVKVDVAHSPOVDLREBELIALGAIIRPAAAVMRSTVGVTAQPELGSVYAD 2294  
 QY 788 HAREVVRPADVAKALHAAGACTFVEVGKSTLLGLVPCMPDAP--ALLASSRAGRDEP 845  
 DB 2295 NLRQVVRPAAAOALLEGCPALFTEMSPHILVPLDEIQTAEGCGAAVSLRRGDER 2354  
 QY 846 ATVLEALGLMAVGLVWAGLFPSCGRRVPLPTPMQREBYWITDKADDAARGRAPG 905  
 DB 2355 ATLLEALGTLMAAGPVVSMALFPAGGRVPLPTPMQHERYWI--EDSVHSGK-- 2406  
 QY 906 AGHDEVEGCAVGGDRRSARLDHP----- 930  
 DB 2407 -----PSLRRLQLNGATDHPILGAPLVSARPAHLMEQALSDERLSYLSHRVH 2457  
 QY 931 -----PPSGREKRYEAADRPRLEIDPFG--VLDBLVL-----RYTE----- 967  
 DB 2458 GEAVLPSAAVYEMALAG-----VDLYGTATLVLEQALERALAVSEGRIVQVALS 2510  
 QY 968 RRAPLGEVEIAY--DAAGLSFNDVQALAGV-----PDDLPGK 1004  
 DB 2511 EBGPRASFOVSREBEGRSM--VRHATGHVCSGSSAVGALKENPMTIORRCBSVLSS 2568  
 QY 1005 PNPPL-----LG-CECAGRIYAVEGVNGVLVGGPVIALSAGAFAT 1045  
 DB 2569 ALYPLNHALDYGCPFCGVGQVWLTGTEGVALRPED-----MASSGAVRI 2617  
 QY 1046 H-----VTTSAALV-----LPP-----QALSAT----- 1064  
 DB 2618 HPALIDACFOVLTALLTTPSEIIRRLRLDLEHBDLPSRPAVNOAVSDTWLMDALDGG 2677

QY 1065 --EAAAMPV-----AYLTAWYALDRIR-----LOGGERVLIHAATGGVGLA 1104  
Db 2678 RROGASVPEVDVLVLSFPAKWEVMERLAQAYIIIGTLRIWNVFCAGERTIDELVRLQIS 2737  
QY 1105 AV-----QNAQH-----VGAEVHATAG-----TPEKRAYLESIGVRYVSDSRDRFVA 1147  
Db 2738 VVYRKVIKRWMEHLVAIGILVGDEHVFSSQPLPEPDIAVLEENG-----RVFA 2787  
QY 1148 DVRA---WT---GGEGVDVVL-----NSLSGELI 1170  
Db 2788 DLPVLFKCKPAGERLADVLTKTLALEILFPGGSFDMARIYRDSPIARYSNGIVRGV 2847  
QY 1171 DKSPNLSRSHGRF---VELG-----KRDGY-----ADNOLGLR 1200  
Db 2848 ESAARVVAPEGMFSLIEIGAGTGAATAVLPVLPDRTEYHFTDVSPLFLARAEGRFRDY 2907  
QY 1201 PFLRNLSFSLVD-----LRGMMLERPARVRLLELLGLIAGVF-- 1240  
Db 2908 PFLK---YGIIDVDOEPAGOGYAHQRPDVIIVANVIHATRDIRATAKRLLSILAPGGILV 2964  
QY 1241 ---TPPI-----ATLPIARVADAFPSMAQAQHL-----GK 1268  
Db 2965 LVEGTGHPIMFDITTTGLIEGWOKYEDDLRIDHPILPARTWCVDLRRVGFADAVSLPGDGS 3024  
QY 1269 LVLTLDGPEVQIRIPTHAGAG-PSTGDRDLDRLASAPAPAAA-----A 1312  
Db 3025 PAGILGGHVIUSRAAGIAGACDSSGES-----ATESPARAVRQEWADGSADVYHRMA 3078  
QY 1313 LEA--FLRTQVSQVLRTPEIVGAEALFTRLGMSLSMAVEILNRIEASLK 1360  
Db 3079 LERNYFHRPGRQYVWHGRRLRTGGGA-FTKALAGDILLFEDTGCQVAEVQ 3127

Search completed: October 2, 2003, 17:33:51  
Job time : 133.763 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 37.1179 Seconds  
(without alignments)  
2912.986 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124

Sequence: 1 MTOEQANOSETKPAFDKFPK.....HPAFNIESLNVILKPSXAG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2124	100.0	419	2	Q9L8C4
2	2107	99.2	419	2	Q9L8C4
3	532.5	25.1	395	2	Q8KZM6
4	497	23.4	417	2	Q8R3Z4
5	488	23.0	416	2	Q87605
6	486.5	22.9	415	16	Q92Q05
7	486	22.9	376	16	Q31785
8	483	22.7	399	2	Q93HA3
9	481.5	22.7	418	16	Q9HYR4
10	469	22.1	419	2	Q8KX32
11	459	21.6	416	2	Q8KX32
12	453.5	21.4	410	2	Q9XSP9
13	451	21.2	415	16	Q8UFY8
14	447.5	21.1	401	2	Q9KHU7
15	446.5	21.0	394	2	Q8KNC8
16	443	20.9	399	2	Q8GGQ1

17	442.5	20.8	434	16	Q9CBE7	Q9CBE7 mycobacteri
18	441	20.8	433	16	Q9F200	Q9F200 streptomyc
19	433.5	20.4	399	2	Q93H81	Q93H81 streptomyc
20	431	20.3	379	2	Q93HFO	Q93HFO streptomyc
21	430	20.2	411	2	Q60005	Q60005 streptomyc
22	430	20.2	411	2	Q59910	Q59910 streptomyc
23	429	20.2	411	16	Q9X8G3	Q9X8G3 streptomyc
24	427.5	20.1	407	16	Q9RJ07	Q9RJ07 streptomyc
25	427.5	20.1	416	2	Q93H80	Q93H80 streptomyc
26	425	20.0	511	2	Q52569	Q52569 amycolatops
27	423	19.9	397	2	Q8KIC8	Q8KIC8 nocardia ae
28	421.5	19.8	444	16	Q91107	Q91107 pseudomonas
29	420	19.8	411	2	Q32460	Q32460 actinomadr
30	417	19.6	407	2	Q59819	Q59819 streptomyc
31	416.5	19.6	398	16	Q8XRX2	Q8XRX2 ralsconia s
32	413	19.4	397	2	Q59523	Q59523 micromonos
33	412	19.4	397	16	Q8PHS9	Q8PHS9 xanthomonas
34	411	19.4	404	2	Q54302	Q54302 streptomyc
35	409.5	19.3	416	16	Q982V9	Q982V9 rhizobium 1
36	408	19.2	404	2	Q93HU0	Q93HU0 streptomyc
37	407.5	19.2	436	2	Q9S4D6	Q9S4D6 streptomyc
38	406	19.1	310	2	Q32927	Q32927 mycobacteri
39	405.5	19.1	418	2	Q8GME6	Q8GME6 streptomyc
40	403.5	19.0	420	2	Q92HQ1	Q92HQ1 streptomyc
41	402.5	19.0	406	2	Q87192	Q87192 streptomyc
42	401.5	18.9	397	2	Q8KND6	Q8KND6 micromonos
43	397.5	18.7	416	3	Q96MS9	Q96MS9 trichosporo
44	396.5	18.7	397	16	Q8P6C3	Q8P6C3 xanthomonas
45	393	18.5	399	2	Q93NKL	Q93NKL streptomyc

## ALIGNMENTS

RESULT 1	Q9L8C4	PRELIMINARY;	PRT;	419 AA.
AC	Q9L8C4;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Cytochrome P450 167A1.			
GN	CYP167A1 OR BPOF.			
OS	Polyangium cellulosum.			
OC	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;			
OC	Scorariaceae; Polyangiaceae; Polyangium.			
OX	NCBI_TaxID=56;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SO CE90;			
RX	MEDLINE=20130945; PubMed=10662695;			
RA	Molnar I., Schnupp T., Ono M., Zirkle R.E., Milamow M.,			
RA	Nowak-Thompson B., Engel N., Toupet C., Strittmann A., Cyr D.D.,			
RA	Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;			
RT	"The biosynthetic gene cluster for the microtubule-stabilizing agents			
RT	epoхиlones A and B from Scoringium cellulosum So ce90."			
RL	Chem. Biol. 7:97-109(2000).			
CC	-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
DR	EMBL; AF210843; AAF26924.1; -.			
DR	HSSP; P33006; ICPT.			
DR	InterPro; IPR001128; Cytochrome_P450.			
DR	Pfam; PF00067; P450; 1.			
DR	PRINTS; PR00385; P450.			
KW	Heme; Monooxygenase; Oxidoreductase.			
SQ	SEQUENCE 419 AA; 46750 MW; 13E5857FB04FB8FE CRC64;			
Query Match	100.0%; Score 2124; DB 2; Length 419;			
Best Local Similarity	100.0%; Pred. No. 8.5e-145;			
Matches	419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 MTOEQANOSETKPAFDKFPAGVADPPPAIERLRKATPIFYWEGSGSWLTRHYDVA 60			
DB	1 MTOEQANOSETKPAFDKFPAGVADPPPAIERLRKATPIFYWEGSGSWLTRHYDVA 60			

```
QY 61 VFRDERFAVSREEMESSAAYSSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
|
|
|
Db 61 VFRDERFAVSREEMESSAAYSSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
|
|
|
QY 121 LRAEIQRTVDQLDARSQGEFDDVRYDAEGIPMRAISALLKVPACDEKFRFGSATA 180
|
|
|
Db 121 LRAEIQRTVDQLDARSQGEFDDVRYDAEGIPMRAISALLKVPACDEKFRFGSATA 180
|
|
|
QY 181 RALGVGLVPOVDEETKTIVASVTEGLALHLDVDERRRNPLENDVLTMLLQAEADGSRUS 240
|
|
|
Db 181 RALGVGLVPOVDEETKTIVASVTEGLALHLDVDERRRNPLENDVLTMLLQAEADGSRUS 240
|
|
|
QY 241 TKELVAVGAIIAAGDTTIIYLAFAVNLRLSPPEALVELKAPGLMRNALDEVLPDNI 300
|
|
|
Db 241 TKELVAVGAIIAAGDTTIIYLAFAVNLRLSPPEALVELKAPGLMRNALDEVLPDNI 300
|
|
|
QY 301 LRIGTVRFARQDLEYCGASIKKGEWFLIPALRDGTFSRDPVDFVDRDTSASLAYS 360
|
|
|
Db 301 LRIGTVRFARQDLEYCGASIKKGEWFLIPALRDGTFSRDPVDFVDRDTSASLAYS 360
|
|
|
QY 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPFRNIESLNVILKPSKAG 419
|
|
|
Db 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPFRNIESLNVILKPSKAG 419
|
|
|
RESULT 2
ID 09K124 PRELIMINARY; PRT; 419 AA.
AC 09K124;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE BPOK.
GN BPOK.
OS Polyangium cellulolum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Soranginae; Polyangiaceae; Polyangium.
NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RA MEDLINE=2029308; PubMed=10831849;
RA Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT "Isolation and characterization of the epoH1 gene biosynthetic gene
cluster from Sorangium cellulolum."
RL Gene 249:153-160 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RX MEDLINE=20115953; PubMed=10649995;
RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT "Cloning and heterologous expression of the epoH1 gene cluster."
RL Science 287:640-644 (2000).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF217189; AAF62886.1; -
DR HSSP: P33006; 1CPT.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450.
DR PRINTS: PR00385; P450.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 419 AA; 46750 MW; FCCEDAI93563688F CRC64;

Query Match 99.2%; Score 2107; DB 2; Length 419;
Best Local Similarity 99.3%; Pred. No. 1,4e-143;
Matches 416; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Db 61 VFRDERFAVSREEMESSAAYSSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
|
|
|
QY 121 LRAEIQRTVDQLDARSQGEFDDVRYDAEGIPMRAISALLKVPACDEKFRFGSATA 180
|
|
|
Db 121 LRAEIQRTVDQLDARSQGEFDDVRYDAEGIPMRAISALLKVPACDEKFRFGSATA 180
|
|
|
QY 181 RALGVGLVPOVDEETKTIVASVTEGLALHLDVDERRRNPLENDVLTMLLQAEADGSRUS 240
|
|
|
Db 181 RALGVGLVPOVDEETKTIVASVTEGLALHLDVDERRRNPLENDVLTMLLQAEADGSRUS 240
|
|
|
QY 241 TKELVAVGAIIAAGDTTIIYLAFAVNLRLSPPEALVELKAPGLMRNALDEVLPDNI 300
|
|
|
Db 241 TKELVAVGAIIAAGDTTIIYLAFAVNLRLSPPEALVELKAPGLMRNALDEVLPDNI 300
|
|
|
QY 301 LRIGTVRFARQDLEYCGASIKKGEWFLIPALRDGTFSRDPVDFVDRDTSASLAYS 360
|
|
|
Db 301 LRIGTVRFARQDLEYCGASIKKGEWFLIPALRDGTFSRDPVDFVDRDTSASLAYS 360
|
|
|
QY 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPFRNIESLNVILKPSKAG 419
|
|
|
Db 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPFRNIESLNVILKPSKAG 419
|
|
|
RESULT 3
ID 08K2M6 PRELIMINARY; PRT; 395 AA.
AC 08K2M6;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 23, Last annotation update)
DE Cytochrome P450 enzyme.
GN BPOI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OK2;
RA Sasaki M., Koriyagi T., Kuruu Y.;
RT "Genetic analysis of biotin operon in Bacillus subtilis natto OK2."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AB088066; BAC03244.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 395 AA; 44869 MW; 6D4B5C445107D225 CRC64;

Query Match 25.1%; Score 532.5; DB 2; Length 395;
Best Local Similarity 32.1%; Pred. No. 3.3e-30;
Matches 130; Conservative 96; Mismatches 138; Indels 41; Gaps 11;

QY 24 YAEPPAERLREATPIFYWDEGR-----SWLTRYHDSAVFRDERFAVSREEMESSA 78
|
|
|
Db 12 FLKNPYSFYDTLRAVAHPY---KGSFLKYPGWVVTYETTAALIKDARFKVPTLPBESST 68
|
|
|
QY 79 EYSSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAIDLRAEIQRTVDQLDARS 138
|
|
|
Db 69 KXQ----DISHVGNQWMLFONQDHRRLRTLSGATTPATBESYQYIETVHQLDDYO 124
|
|
|
QY 139 GQEPDVVDVYAGIPMRAISALLKVPACDEKFRFGSATAALGVGVPOVDEETKTIL 198
|
|
|
Db 125 GKKMVEISDFAPPLASPIVANTIGVPEEDRQLKMAAS-----LIQTIDFRSRK 176
|
|
|
QY 199 VASVTEG-----LALHLDVDERRRNPLENDVLTMLLQAEADGSRUSLTKELVALVGA 251
|
|
|
Db 177 V--LTEGNHVAQAMAVYFKELIQKRKH--QQMISMILKKG--ENDKLTDEEASATCIIL 232
|
|
|
QY 252 IAAAGDTTIIYLAFAVNLRLSPPEALVELKAPGLMRNALDEVLPDNIIRIGTVRFAR 311
|
|
|
Db 233 AINGHETVNLISNVICLQHPDQILKRENDLIGTAVEBCLRYESPTQM--TARVASE 291
|
|
|
```







QY 126 ICRVTDQLDARSQGEFVDVYAEIGIPMAISALLKVPACDEKFRFGSATAAL-- 183  
DB 124 IEQIAEQLDDEMEKEKADIMKSFASPLPIVISELMGPKEDRSQFOIWTNAMDTSRG 183  
QY 164 GUGLVPQVDEKTLVASTVEGLALLHDVLDERRRNPLENDVLTMLQEAAGSRISTGE 243  
DB 164 NRELNVQALREFKDYIAK-----LTHD---RRKP-KDDLISKVHAENGSKLSEKE 232  
QY 244 LVALGATIAAGTDTTIIYIAFAVNLILRSPEALIELVKAEPGLMRNALDEVLPFNILRI 303  
DB 223 LYSMLFLVAVLETTVNLGSGTALLQHKKECEKQKQPMIATVAVELLRIYSPVM 292  
QY 304 GTVPARQDLEYCGASIKKGEVFLIIPALRDGVFSPPVDVARDTGASLAYGRCPH 363  
DB 223 MANRMAIEDFTYKSGIKRGMWIFIGISANDPNEFENPEILINRSPNRIISFGFIH 352  
QY 364 VCPGVSLARLEAIVGTIF 383  
DB 353 FCLGAPLARLEGHIAFKAAF 372

## RESULT 8

QY 093HA3 PRELIMINARY; PRT; 399 AA.  
AC 093HA3;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Cytochrome P450.  
OS Streptomyces avermectilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2177403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermectilis: Deducing the ability of producing secondary  
RT metabolites."  
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; AB070947; BAB6286.1; -.  
DR HSSP; P23295; IJFB.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; P450; 1.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Heme; Monooxygenase; Oxidoreductase.  
SQ SEQUENCE 399 AA; 44538 MW; CBOD90378008F99D CRC64;

Query Match 22.7%; Score 483; DB 2; Length 399;  
Best Local Similarity 32.2%; Pred. No. 1.2e-26;  
Matches 126; Conservative 66; Mismatches 163; Indels 36; Gaps 7;

QY 24 YAEDEPPIERLEATPIFYWDEGR---SWVLRHYDVSAVFRDRFAVSRE---EMES 76  
DB 3 FGADYPERAMLRABEPRVQVLEGRGLVGLVTRIEDVRKLSIDPRMSADPNNAFLDWOE 62  
QY 77 SAEYSSAIPELSDMKKYG--LFGLPPEDEHARVRLKVNPSFTSRAIDLRAEIORTVDQL 134  
DB 63 AGKGR-----PLEDRDGLGTHLLTDAPEHTRRLRLVSTAFARVGLRAQVQHTDGL 118  
QY 135 DARSQGEFVDVYAEIGIPMAISALLKVPACDEKFRFGSATAALGVLVQVDEE 194  
DB 119 DTIVRGQAEILGDFAPLAITVICELLGVPKADQVFRQWTKDFFR-----W 166  
QY 195 TKTLVASVTEG-----LALHDVLDERRRNPLENDVLTMLQEAAGSRISTREL 244  
DB 167 TMTDSAQDGRGARGVGLRDLLEYLTRLVDRKRPDAG-LVDALLIARDDDRINAEAL 225  
QY 245 VALVGAIIAAGTDTTIIYIAFAVNLILRSPEALIELVKAEPGLMRNALDEVLPFNILRIG 304

DB 226 LSMMSLLVGFETFTVNLIGNGTALLLRHPQALALRRRPELVDSALEMRLYDGSFETA 285  
QY 305 TVRPARQDLEYCGASIKKGEVFLIIPALRDGVFSPPVDVARDTGASLAYGRCPH 364  
DB 266 TWRFPLEPIEAVAGRIEKGHPVLLSLASANDGAKFAPDPDVTTRADPAHVAFGRAHF 345  
QY 365 CPGVSLARLEAIVGTIFRRFPPEMKLKEPT 395  
DB 346 CUGAPLARLEGRIRAHGLRLRLPGIALSVPP 376

## RESULT 9

QY 09HYR4 PRELIMINARY; PRT; 418 AA.  
AC 09HYR4;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Putative cytochrome P450 107S1.  
GN CYP107S1 OR PA3331.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stever C.K., Pham X.-Q.T., Erwin A.T., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Wu A., Iarbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Kas A., Paulsen I.T.,  
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RT Nature 406:959-964(2000).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; AE004755; AAG06719.1; -.  
DR HSSP; Q00441; IOXA.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR Pfam; PF00067; P450; 1.  
KW Heme; Hypothetical protein; Monooxygenase; Oxidoreductase;  
KW Complete proteome.  
SQ SEQUENCE 418 AA; 46377 MW; 8231E91DB10F351 CRC64;

Query Match 22.7%; Score 481.5; DB 16; Length 418;  
Best Local Similarity 33.5%; Pred. No. 1.7e-26;  
Matches 116; Conservative 73; Mismatches 140; Indels 17; Gaps 5;

QY 50 WVLRHYDVSAVFRDRFAVSREMESSAEYS---SAIPELSDMKKYGGLGPPEDHAR 105  
DB 41 WVVRIRYRARKVRLNHP--GVRRDARQAELVAKRTGSPRAGIGELSHHMLNLPDHDTR 98  
QY 106 VRKLVNPSFTSRAIDLRAEIORTVDQLDARSQGEFVDVYAEIGIPMAISALLKVP 165  
DB 99 LRSLVGRAFTROVERLDPHIERITTEALDLMAGREQDLADPAIPLTIIVIFELGIP 158  
QY 166 AECEKFRFGSATAARALGVLVQVDEETTLVASTVEGLALLHDVLDERRRNPLENDV 225  
DB 159 EAEREHARQSMERQAEILLS-----PEAQALDAQVDYRVL---LEAKRRPAD-DV 207  
QY 226 LTMLOEAAGDSRISTRELVALVGAIIAAGTDTTIIYIAFAVNLILRSPEALIELVKAEPG 285  
DB 208 YSGLVQAADDESGQISEALVSMALHMSGPEITTMNIGNALVTLVNPEDQALALRQPE 267  
QY 286 LMRNALDEVLPFNILRIGTRFAPARQDLEYCGASIKKGEVFLIIPALRDGVFSPPVDV 345  
DB 268 LIPNMAEELVNHDPVARSMTKRTVEDEVLDGVITIPAGEYILVSNLTNHNHAAERFDDPR 327  
QY 346 FVVRDGTGASLAYGRGPHVCPGVSLARLEAIVGTIFRRFPPEMKL 391

DB 328 LDIIRNTDGLGYGFGVHYCVGASLARLEGRALQRLARFPDQL 373

RESULT 10

Q8KY32 PRELIMINARY; PRT; 419 AA.

AC Q8KY32; 01-OCT-2002 (T-EMBLrel. 22, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Rubu.

GN RUBU.

OS Streptomyces collinus.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomycetes.

OK NCBI\_TaxId=42684;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=DSM2012;

RA Saito H., Bruenker P., Martin R., Minas W.;

RT "Streptomyces collinus DSM2012 rubromycin biosynthesis gene cluster.";

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; AF293355; AA097370.1; -

DR InterPro; IPR001128; Cytochrome\_P450.

DR Pfam; PF00067; P450; 1.

DR PROSITE; PS00086; CYTOCHROME\_P450; 1.

KW Heme; Monooxygenase; Oxidoreductase.

KM SEQUENCE 419 AA; 46331 MW; 361266AAFE79DE5B CRC64;

SQ

Query Match 22.1%; Score 469; DB 2; Length 419;

Best local Similarity 30.6%; Pred. No. 1,3e-25;

Matches 121; Conservative 72; Mismatches 169; Indels 34; Gaps 6;

DB 20 FADGVAEDPPAIERLRE--ATPIFYWDEGRSVLTTRYHDVSAVFERDERFA---VSRE 72

DB 16 FTFAFQNPHEALAGLRRTAPAVPMTPNGLRTMLVTGHEBARALLADPRLSKDMVGRD 75

DB 73 -----EMESSAAYSSAIPELSDMKYKGLFGLRPEDHARVRKLVNSFTSRAL 119

DB 76 LIPRNFVDPDKQREFLAESGERSQFPHVLSVH--MLSDSPDHTLRRLRVGSAFTARRV 132

DB 120 DLIRAEIQTVDQLDARSQGEFVDVRYDAEGIPMAAISALLKVPACSEKERRFGSAT 179

DB 133 ESARPPITELTDLLDMARHERLDLMEALAFVPVTVICMLGVPDDDAARRRNSNL 192

DB 180 ARALGVGLVQVDEETKTLVASVTEGLALHLDVLDERRRNPLENDVLTMLLQADGSR 239

DB 193 VSGAG--TDEVRASASMTYLT-----LIEAKRNEPADDMLTDLVHARDAGDOL 241

DB 240 STEELVALVGAIIAGDITIVLIAFAVNLILSPLELVKAEFGLMRNALDEVLRDN 299

DB 242 SSDELLSMALFLVAGHETVNLINGALALLTHPEVEQLADDESIMPQAVEEFLKYDG 301

DB 300 ILRIGTVRFARODLEYGASIKKGEWFLILPSALRDGVFSRDPVDFVARDGASLAVG 359

DB 302 PVTNATMRFTPEVGVSVTIPBGEFVTISIGAAGRPDXYPPDRDLITRAHSGVAFG 361

DB 360 RGPVCPGVSLARLEAIVAGTIFRRPEMKLKEP 395

DB 362 HGHHICIGAPLARLEGRIVLSRLFARLPGRLAADP 397

RESULT 11

Q8KSYO PRELIMINARY; PRT; 416 AA.

AC Q8KSYO; 01-OCT-2002 (T-EMBLrel. 22, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Putative cytochrome P450 Gth03.

GN GTH03.

OS Streptomyces sp. JP95.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomycetes.

OK NCBI\_TaxId=195040;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=JP95;

RA Li A., Piel J.;

RT "A gene cluster from a marine Streptomyces encoding the biosynthesis of the aromatic spiroketal polyketol griseorhodin A.";

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; AF509565; AA33670.1; -

DR InterPro; IPR001128; Cytochrome\_P450.

DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME\_P450; 1.

KW Heme; Monooxygenase; Oxidoreductase.

KM SEQUENCE 416 AA; 46093 MW; 4E33B40A11EBD4D9 CRC64;

SQ

Query Match 21.6%; Score 459; DB 2; Length 416;

Best local Similarity 31.8%; Pred. No. 6,9e-25;

Matches 134; Conservative 76; Mismatches 154; Indels 58; Gaps 13;

DB 3 QEOANQSETKPAFDK--PF--ARGYAEDPPAIERLREATPI--FYWDEGRSVLT- 53

DB 9 QTLADGAEDAPAVPVYKTCPPYRMPGY-----EELREKGPISRVTLMNGRTAMLVYG 60

DB 54 -----RYHD--VSANFDERFAVSRREWESSAAYSSAIPELSDMKYKGLFGLRPEDHAR 105

DB 61 NDIGRRLLFPDARLSSVDLPFRPLAPRIEAGRQQAADP-----LVGDDEVHAR 111

DB 106 VRKLNVPSFTSRALDILRAEIQRTVDQLD--ARSQGEFVDVRYDAEGIPMAISALL 162

DB 112 QRMTVPSGICQINLRPEIQKXADDDLTMLAKPGVTVLITREYALPMNSAVICML 171

DB 163 KVPAAE---CDEKFRFGSATARALGVGLVQVDEETKTLVASVTEGLALHLDVLDERR 218

DB 172 GVPYEDHHYFDRSRRLVSSG-----EEQAAQQAQFTEILAYLDDLIVRKQA 220

DB 219 NPLENDVLTMLLQAEADGSRSTKELVALVGAIIAGDITIVLIAFAVNLILSPLE 278

DB 221 EP-GDTLDLDELARQLEBEGKVDROELAMATVTLVSGHETTNMIALSTMLLADDPQLA 279

DB 279 LVKAEFGLMRNALDEVLRDNILRIGTVRFARODLEYGASIKKGEWFLILPSALRDGT 338

DB 280 ALRADESLMPRAVDLMRFTSSIGDM-LMRVAKEDIIEBHLLRAGGVILSTMLMNRDPG 338

DB 339 VFSRDPVDFVRDGTGASLAVGGRPHVCPGVSLARLEAIVAGTIFRRPEMKL---KET 394

DB 339 AFERPDELDIRRPAGRHVAFGYIHQICQNLARAMEIATLTLFRFVPTLKLAVAPEQV 398

DB 395 PV 396

DB 399 PV 400

RESULT 12

Q9XSP9 PRELIMINARY; PRT; 410 AA.

AC Q9XSP9; 01-NOV-1999 (T-EMBLrel. 12, Created)

DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Putative cytochrome P450 107N1.

GN CYP107N1.

OS Streptomyces lavendulae.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomycetes.

OK NCBI\_TaxId=1914;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL 2564;

RX MEDLINE=99201491; PubMed=10099135;

RESULT 14	
Q9KHJ7	
ID	Q9KHJ7
AC	Q9KHJ7;
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	Putative cytochrome P450 107R1 (Putative cytochrome P450 monooxygenase Enclr).
DE	
GN	CYP107R1 OR ENCR.
OS	Streptomyces maritimus.
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC	Streptomycinae; Streptomycetaceae; Streptomyces.
OX	NCBI_TaxId=115828;
RN	[1]
RP	SEQUENCE FROM N.A.

RA Piel J, Hoang K., Moore B.S.;  
RL "Natural Metabolic Diversity Encoded by the Enterocin Biosynthesis  
RT Gene Cluster";  
J. Am. Chem. Soc. 122:5415-5416(2000).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; AF254925; AAF81737.1; -.  
DR HSSP; 000441; IOXA.  
DR InterPro: IPR001128; Cytochrome\_P450.  
DR Pfam: PF00067; p450\_1.  
DR PRINTS; PR00385; P450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Heme; Monooxygenase; Oxidoreductase.  
SQ SEQUENCE 401 AA; 44184 MW; 73CDA22BD475B2E7 CRC64;

Query Match 21.1%; Score 447.5; DB 2; Length 401;  
Best Local Similarity 31.9%; Pred. Misat. 4.4e-24;  
Matches 126; Conservative 70; No. matches 166; Indels 33; Gaps 12

QY 9 SETKPAFPKKPAPGVAEDPPFAISRLLRATPIF-----YMDEGRSWLTTRYHDVSAVR 63  
DB 3 THTQQLRDF-PAPAELHMEBAFAQLNREEEISKRLEPY--GGEMALVTRIQDIKTIVLG 59  
QY 64 DERFAVSREWESESSAESAIPELSDMKKYG-LFGIPEPDHARVRKLVPSETSPAIDL 122  
DB 60 DPRFS-----RAAQHQAPRIQDPDAGEGLVMSIDPDHTRLRKTAVGVFTKGRVEDL 113  
QY 123 RAETIRTYDQLIDA-RSGOEEDVDVRYDAEGIPMAISAALLKVPAECDEKFFRFSGATKR 181  
DB 114 RPATRIAIEELLEAMEBASCAPDLVASVALPLPVTVICDLLGVPGDDRBQLRGWSDAL-- 171  
QY 182 ALGVGLVNPQVDEBTTLVASVEGIALTHDVDERRRRNLENVDVITMLQA-EADGSRLS 240  
DB 172 ---ISTTACTPESAAQAAMDHFRAL--VSQRKPPTD-DLIGALVQTMDREGILR 224  
QY 241 TKELVALVGAIITAAGDTTIYLIAFAVLNLIRSPEALELVKAEPGIMRNALDEVIRFDNI 300  
DB 225 DELVILTRFDLLIAGHETTASQIANCTYLLQRPHMDRLRTDPSMASAVBELRLF--- 281  
QY 301 LRIGT---VRARODLEYCGASIKKGENVFLILSALRDGYVFSPDYVDVRRDTGSL 356  
DB 282 IPLGGSFPARAVATEPVELCGVRIOPGDVFPPTVAANMDDPVFAEPGRILDISPENHY 341  
QY 357 AYGRGPVPCGVSLARLEAEIVGTFRRFPFMKL 391  
DB 342 AFGHGHHCLGAQLRLQLVALGLVLRRLPRURL 376

RESULT 15  
Q8KNK8 PRELIMINARY; PRT; 394 AA.

ID Q8KNK8 AC Q8KNK8; PRT; 394 AA.  
AC Q8KNK8; PRT; 394 AA.  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE CALB10.  
GN CALB10.  
OS Microspora echinospora (Microsporida purpurea).  
OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;  
OC Micromonosporineae; Micromonosporaceae; Micromonospora.  
OX NCBI\_TaxID=1877;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRL 15839.  
RA Ahlert J., Shepard E., Lomovskaya N., Zazopoulos E., Staffa A.,  
Baumann B.O., Huang K., Feinstein U., Czişny A., Miltwan R.E.,  
Parret C.M., Thorson J.S.;  
RT "The calicheamicin gene cluster and its iterative type I PKS.";  
RL Science 0:0-0(2002).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
DR EMBL; AF497482; AAM94800.1; -.  
DR InterPro: IPR001128; Cytochrome\_P450.  
DR Pfam: PF00067; p450\_1.  
DR PRINTS; PR00385; P450.

```

KW PROSITE: PS00086; CYTOCHROME P450; 1.
DR Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 394 AA; 43871 MW; AA79A2518A2EFBDC CRC64;

Query Match      21.0%; Score 446.5; DB 2; Length 394;
Best Local Similarity 32.1%; Pred. No. 5,16-24;
Matches 131; Conservative 64; Mismatches 162; Indels 51; Gaps 13.

Oy 19 PEAQGYADEPPRA-IERLRREATPIF-----YMDGRGSWLTRYHDVASVPRDERFAVSRE-- 72
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 PFGR-----PAFYARLRTERPVARLRPLMLGNTAMVVSRYADVKRYLSDPRMSADRRA 58
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 73 -----EWSSAEYSSAIPELSDMKYGLFGIIPEDHARVKKLVNPSFTSAIDLRL 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 GPFPRAPITBESQGRQASFRANFRPLNMND-----PPE-HTAARQIYDEFAARVRQLR 110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 124 AEIORTVDQLLDA-RSQGEFEDVVRDYAEGIIPKRAISALKVPAECDEKFFRFGSATARA 182
    ::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 111 PLVERVVDDELHDADTAGRSSADLVPSFSYPVPSRVTCEMLGVPGYEHAFER---RSTRM 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 183 LGVGLVQVQVDEETTLVAASVTEGLALHLDVDERRRNPLENDVLTMLLQAADSRLSTK 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 LSRG-VP-ADBRAR---CAREIREFLDGVGVTDEKRP-GDGVLSRLLAAPRAAGEPDHE 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 243 ELVALVGAIIAGADTTIYLAFAPVNLNRSPELMEIYKAPGGLMRNALDLEVLRFNILR 302
    ::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 221 AVSMAMFVLVAGVHTTSMNLSVSLVALLTHPEKLARLRAPDRFPRAVEELLRTFIVE 280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 303 IGVTRFARQDLEYCGASIKGEMVFLIPSLARDGVTSRPDVFVRDRDTGASLAYGGRP 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 AATARTATADVTVGVTTRBAGEGVVALGQANRPDAFDRDEDFDRDADAHHLAFGYGR 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 363 HVCGGVSLARLEALANGTIRRRPEMKL-----KETPVGYN 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 HICGGHILARLELDVALSRIVRRPGRLLTVDVDDLPLKEGNIFFGLH 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: October 2, 2003, 17:31:48
Job time : 41.1179 secs

```



```

QY 130 VQDLDRSGQEPFVVRDYABGIPMRATISALKVPAECDEKRRRGATATRALGIVLP 189
DB 128 TDLLDRLRGLRSSVDLIDEPAPLPTIVSELGVDSRRDRD---TNLVGSGSP 183
QY 190 QVDEETK-TLVASVTEGLALHDVDERRRNPLENDVLTMLQADGSRSTKELVALY 248
DB 184 EAOQASVAMVEVLTETLIA-----KRTGPDOLLTRLEAVEDGRLSGELIANV 235
QY 249 GAIIAAGDTTTLIAFAVLNLRSPALVELVKAEPGLKRNALDEVLRFDNIIRIGTVPR 308
DB 236 FLLLVAGHETTVLINGCVSLILGNPDQLALRNDPSLLFGAIEETLRYESPVANGTFPH 295
QY 309 ARODLEYCASIKKGMVFLIIPSAIRDGVFSRPDVPDRRTGASLAVGRGPHVCPGV 368
DB 296 TAEAVFPGDVVPEGLVWALGANRRDGERFEDPDRFDTTRRTTGHVAGHGIFCVGA 355
QY 369 SLARLAEIAGVTIFRRPEMKLETP---VFGYHAFRNIESLNV 411
DB 356 ALARLEAQIAVGRLLRFRPDLRMAASPDRLRMFRFVLMRGLEKLPV 401

RESULT 2
BIOI_BACSU STANDARD; PRT; 395 AA.
AC P53554;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin biosynthesis; cytochrome P450-like enzyme (EC 1.14.-.-).
OS BIOI OR CYP107H.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96312354; PubMed=8763940;
RA Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahaim P.,
RT "Cloning, sequencing, and characterization of the Bacillus subtilis
RT biotin biosynthetic operon."
RT J. Bacteriol. 178:4122-4130(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rrm-dnaB region."
RT Microbiology 143:3431-3441(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Batters M.G., Bessieres P., Bolochin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano N.J., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Danisco F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Eutenei K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritsch C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaeser P., Goffeau A., Golligly E.J., Grandi G.,
RA Hilsenrath G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasehara Y., Kleier-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Portollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

```

```

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
RA Sekiguchi U., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Toseco V., Uchiyama S., Vandenbol M., Vanlier F., Vassariotti A.,
RA Viari A., Wambat R., Wedler E., Wedler H., Weltzegeger T.,
RA Winere P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U51868; AAB17462.1; -
CC DR EMBL; AF008220; AAC00266.1; -
CC DR EMBL; Z99119; CAB14997.1; -
CC DR PIR; G69594; G69594.
CC DR HSSP; Q55080; 1107.
CC DR Subtilisin; BG11528; biot.
CC DR Interpro; IPR001128; Cytochrome_p450.
CC DR Pfam; PF00067; P450; 1.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KM Biotin biosynthesis; Oxidoreductase; Monooxygenase;
CC KM Electron transport; Heme; Complete proteome.
CC FT METAL 345 345 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SQ SEQUENCE 395 AA; 44865 MW; E4AC3AF2637ACE1A CRC64;

Query Match 24.7%; Score 525.5; DB 1; Length 395;
Best Local Similarity 31.9%; Pred. No. 6; 5e-26;
Matches 129; Conservative 98; Mismatches 137; Indels 41; Gaps 12;

QY 24 YAEPPPAERLREATPIFYWDEGR-----SWVLTRYHDVSAVFRDERFAVSREMESSA 78
DB 12 FLKNFYSFYDTRLRAVPIY---KGSFLKYPGVWVTCYEETALILKQARFKVTRPLPESST 68
QY 79 EYSSAIPELSDMKKTYLFGLPREDHARVKYNPSTSAIDLRLAEIORTQDOLLARS 138
DB 69 KYQ---DLSHVQNMMLFQONOPDHRRLKTLASGAFPTPTTSYQYIETVHHLDDQVQ 124
QY 139 GOEFPVVRDYABGIPMRATISALKVPAECDEKRRFGSATRALGVGLVPOVDEETKTL 198
DB 125 GKKEKEVIDPAPFLASPIYANIGVPEEDRQLKEMAS-----LIQITD-PTRRR 175
QY 199 VASVTEG-----LALHDVDERRRNPLENDVLTMLQADGSRSTKELVALVGI 251
DB 176 KA-LTEGINAMAVQAMVFKELIOKRKRP-QODMISMLKGR-EKKLTREEBAASTCIIL 232
QY 252 IAAAGDTTTLIAFAVLNLRSPALVELVKAEPGLKRNALDEVLRFDNIIRIGTVPRFQ 311
DB 233 AIAAGHTTNLINSVLCILQHEQLKREPNDLGTVERGLRESPTQM-TARVAS 291
QY 312 DLEYGASIKKGMVFLIIPSAIRDGVFSRPDVPDRRTGASLAVGRGPHVCPGVSLA 371
DB 292 DIDICVTRKQGOVLVLGGANRRDSITNPDPVPTITSPPNHLSPFGHGHVCGLOSSLA 351
QY 372 RLEAEIAGVTIFRRPEMKLETP---VFGYHAFRNIESLNV 411
DB 352 RLEAQIAINTLLQRMPSINLADPWRVRLFG---FRALIELPV 392

RESULT 3
YUIB_BACSU STANDARD; PRT; 396 AA.

```



AC 034374;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putative cytochrome P450 yj1B (EC 1.14.-.-).  
 GN yj1B  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RC RIVOLTA C., SOLDO B., LAZAREVIC V., JORIS B., MAUEL C., KARAMATA D.,  
 RT "A 35.7 kb DNA fragment from *Bacillus subtilis* chromosome containing a  
 RT putative 12.3 kb operon involved in hexuronate catabolism and a  
 RT perfect catabolite-responsive element."  
 RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RC MEDLINE=98044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertiero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fretz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Chim S.Y., Glaier P., Goffeau A., Golligly E.J., Grandi G.,  
 RA Guelsepti G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaat A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaeger-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Portwillik S., Prescott A.M.,  
 RA Presacian E., Pulic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sataie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,  
 RA Sekiguchi J., Sekowska A., Serr S.J., Serr P., Shin B., Soldo B.,  
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Takakoshi A., Tanaka T., Terstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarelli A.,  
 RA Viati A., Wambut R., Wedler E., Wedler H., Weizenecker T.,  
 RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*."  
 RT Nature 390:249-256 (1997).  
 RL -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, AF015825; AAC6317.1; -  
 CC EMBL, Z99110; CAB13078.1; -  
 CC PIR, B69851; B69851.  
 CC HSP, Q55060; 1109.  
 CC Subtilist, BG1195; yj1B.  
 CC InterPro, IPR001128; Cytochrome\_P450.  
 CC Pfam, PF00067; P450; 2.  
 CC PRINTS, PRO0385; P450.  
 CC PROSITE, PS00086; CYTOCHROME\_P450; 1.  
 KM Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
 KM Complete proteome.

FT METAL 349 349 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 396 AA; 44990 MW; 9A89CF12613DBCFB CRC64;  
 Query Match 24.4%; Score 518.5; DB 1; Length 396;  
 Best Local Similarity 32.6%; Pred. No. 1.8e-25;  
 Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;  
 QY 25 AEDPPALERLEATPIPFYWDGSGWVLTTRYDVAVPADEPFAVSREBMESSAEXSSAI 84  
 DB 23 AHFPFWESMKDAPVSGFDEENQWVSFLVDYVKVVGDKL-----FSSGM 70  
 QY 85 PELSDMKYGLFGLPEPDHARVKLVNPSFTSRAIDLRAEIQRYVDLLDRSGQEDD 144  
 DB 71 POTSIGNSINMDPPKTKTSVYVNAFFPRVWKOWEPRIOEITDELIOFGSGSEPD 130  
 QY 145 VVRDVAEGIPMAISALIKVPAECQEKRRFSGATAPALGVGLVQVDEETLVA---- 200  
 DB 131 LVHDSYPLPVIVISELGVPSAHMEQFASD-----LVSTPKDSEAKAFLEED 185  
 QY 201 SYTEGLALHDVLDERRRNPLENDVLTMLQAADGSRSLSTKELVALVGAIIAGDTTI 260  
 DB 166 KCEELAAFPAGIIEKKNKPEQDIISLVEAEETGKLSGSELLPFTLLVAGNETT 245  
 QY 261 YLIAFAVNLRSPPALDELVAEPGLKRNALDEVLPF--DNILRIGTVRFARQLEYCG 317  
 DB 246 NLISNAMYSLTEPPVYEELSHPELMPQAVEALRFAPAPVLR----RIKRPTEICG 301  
 QY 318 ASIKKGEVFLILPBALDGTFSRDPVDPVDRDGSASLARGHVCGLSLARLENI 377  
 DB 302 HIKKGDVLAFAVSADEAKFDRPHMFDIRRHPHIAFGHIFLCGLAPLARLEANI 361  
 QY 378 AVGITFRFPPEMK-LKETPV 396  
 DB 362 ALTSLSAPFPHMECVSITPI 381  
 RESULT 4  
 ID CPXY\_BACSU STANDARD; PRT; 410 AA.  
 AC 008469;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 (EC 1.14.-.-).  
 GN CYP4 OR CYP107J1.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE=97431495; PubMed=9287000;  
 RA Belitsky B.R., Gustafsson M.C.U., Sonenshein A.L., von Wechenfeldt C.,  
 RT "An lrp-like gene of *Bacillus subtilis* involved in branched-chain  
 RT amino acid transport."  
 RT J. Bacteriol. 179:5448-5457 (1997).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97453479; PubMed=9308178;  
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,  
 RA Dusterhoft A., Ehrlich S.D.,  
 RT "Sequence of the *Bacillus subtilis* genome region in the vicinity of  
 RT the lrp operon reveals two new extracytoplasmic function RNA  
 RT polymerase sigma factors Sigv and Sigz."  
 RT Microbiology 143:2939-2943 (1997).  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertiero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,



```

QY 190 QVDEETKTLVASVTEGIALHLDVLDERRRNPENLDVLTMLQADGSRISTKELVALVG 249
DB 187 ---EEVRAADELLEYLARL---ARTKRERPDAAISRLVARGELDTQATNGRLLV- 239
QY 250 AIIAAGTDTTIIYIAFAVINLILSPALVELVKAEPGLMKNALDEVLPFNILIRIGTVRA 309
DB 240 ---AGHETTANMTALSTVLNPNQRLRAEPALVKGAVBELRLRYLTIVNGVPRIA 295
QY 310 RODLEYCGASIKKGEVWFLILIPGALRDGVFSRPDVFDRDGTGASLAVGRGPHVCGVS 369
DB 296 TEDVLIGRTIAGAEVGLCKISSANDAEVFGCDLDDVARDRRRHVAFGVHQLGQP 355
QY 370 LARLEAEIAGTIFRRFPPEKLT---KETP-----VFGYH 400
DB 356 LARVELQIAIETLLRLPLRLAVPHEIIPFGDMALYGVH 396

RESULT 6
C123 MYCTU STANDARD; PRT; 402 AA.
AC P77902;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 123 (EC 1.14.-.-).
GN CYP123 OR RV0766C OR MT0790 OR MTCY369.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxId=1773;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sultson J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RT Nature 393:537-544 (1998).
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L., Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A., Bristel W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z80226; CAB02396.1; -
CC EMBL; AE006970; AAK45032.1; -
CC PIR; A70707; A70707.
CC HSSP; 000441; 10XA.
CC TIGR; MT0790; -.
CC Tubercule; RV0766C; -.

```

```

DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 350 350 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 402 AA; 45421 MW; 76AD843019361798 CRC64;

Query Match 21.2%; Score 450; DB 1; Length 402;
Best Local Similarity 32.4%; Pred. No. 3,4e-21;
Matches 128; Conservative 64; Mismatches 179; Indels 24; Gaps 9;

QY 13 PADEKFPAGYAEADPPFAIERLRREATPIFYWDEGRSWLTRYHDSAVAFRD-----ERF 67
DB 8 PELVADPYVDHEDPPYRYRLRDBAPLYRNEENFMAVSHHHVLDQGFRTSLSNAY 67
QY 68 AVSREWESSAAYSAIPELSPMKYGLFGLPEPDHARVYKLVNDSFSAIDLLRAEIQ 127
DB 68 GVS---LPPSSRTSAYRVMS-----MLAMDPAHLRRTLVSKGFTFRIRLEPQVL 118
QY 128 RTVDQLDARSQGEEDVDVRYAEGI PMRAISALALKVPAECDEKRRPFSATRA -RALGV 186
DB 119 ELARIHLDSALQTESPDEFAEFAGKLPMDVISELIGVPTDRARIRALADAVLHREDVA 178
QY 187 LVPQVDEETKTLVASVTEGIALHLDVLDERRRNPENLDVLTMLQADGSRISTKELVA 246
DB 179 DVP-----PPMAAASII-ELMRYADLIAEFRRRP-ANNITSLALAAELDGDRLSDEIMA 231
QY 247 LVGAIIAAGTDTTIIYIAFAVINLILSPALVELVKAEPGLMKNALDEVLPFNILIRIGTV 306
DB 232 FLFLWVIAIGNETTTLKLANAVYMAAHHPQQLARVPADSHRI PMWVEETLRVYDTSQI -LA 290
QY 307 RPARDLEYCGASIKKGEVWFLILIPGALRDGVFSRPDVFDRDGTGASL -AVGCPHYC 365
DB 291 RTVAHDLTLYDTTIEGEVLLLPSSANRDDRVPDDPDYRIGRIGKLVFSGGAHPC 350
QY 366 PGVSLARLEAEIAGTIFRRFPPEKLTETPVGVYH 400
DB 351 LGAHLARAEARVALGALRLRIRNYEVDDNVYRVH 385

RESULT 7
C140 MYCTU STANDARD; PRT; 438 AA.
AC 008454;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 140 (EC 1.14.-.-).
GN CYP140 OR RV1880C OR MT1929 OR MTCY180.38.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxId=1773;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sultson J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RT Nature 393:537-544 (1998).
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```

RA Frieschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwin M.L., Hatt D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salicberg S.L.,  
 RA Delcher A., Ueberback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 CC Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; 297193; CAB10066.1; -  
 DR EMBL; AE007049; AAK46201.1; ALT\_INIT.  
 DR PIR; E70515; E70515.  
 DR HSSP; Q00441; 10XA.  
 DR TIGR; M11929; -.  
 DR Tuberculist; Rv1880c; -.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KM Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
 KW Complete proteome.  
 FT METAL 381 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 438 AA; 48871 MW; 0463B3EFCF58744 CRC64;  
 Query Match 20.5%; Score 434.5; DB 1; Length 438;  
 Best Local Similarity 29.6%; Pred. No. 3.5e-20;  
 Matches 126; Conservative 77; Mismatches 168; Indels 55; Gaps 13;  
 QY 22 PGYADDFPAIERLREATPIFYWDEGRSWLTRYHDS-AVFRDERFAVREWESSAEY 80  
 DB 35 PAVATDPVPYDEVRSGALV---RNPANLYTDHRLAHLNLSDDRRV-----VSF 83  
 QY 81 SSAIP-----ELSDMKKYGFLPGLEPDHARVKLVNPSFSTSAIDLRAEIQ 127  
 DB 84 GENLPPLRLWLRRTRGDQLPLREPSLAVPEPDHTRRYKRTVASVFTSRVAVSALRDIVE 143  
 QY 128 RTVDQLDARSQGEF-DVVRDYAEGIPKRAISALLKVPACDEKFRFGSATARALGVG 186  
 DB 144 QTAIINLDRFAEQGIIVDVGRYCSQLPVIVISITLGVPEHDPRLVLEFGLAAPSIDIG 203  
 QY 187 LVQVDEETKTLVASVTEGL---ALHADVLDERRRNPLENDVLTMLQAEADG---SRL 239  
 DB 204 I-----PMROYLRVQGGIRGPDWCLEGLHQQARHNP-GDILMSQLIQIABSGDNNTQL 255  
 QY 240 STELVALVGAIIAAGTDTTITVLIAPAVLNLSPEALELVKAEPLGRNALDEVIRFDN 299  
 DB 256 DETELRAIAGLVAVAGFETVNLGNGIRMLDTPBEHLATLRQHPBLMPVTEIIRLDS 315  
 QY 300 ILAIGTRFARODLEYGASIKKGEWVFLIPSLRGCTVPSRPDVDR-DTGASLAY 358  
 DB 316 PVQL-TARVACRQDEAVGVRIKGEVVVYILAAANRPDAVFPDHRDIRPNAGRHLAF 374  
 QY 359 GRGPHVCPGVSRLAEIAVGTIFRRFPEMKL-----KETPVFGYHAFRNIESLNI 412  
 DB 375 STGRHFLGALARAEGEVGLRTFFDFPVRRAAGASRDTRV-----LRGMSITLPT 428  
 QY 413 LKPSKA 418  
 DB 429 LGPARS 434

RESULT 8  
 CPXM\_BACSU STANDARD; PRT; 405 AA.

AC P27632;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 109 (EC 1.14.-.-) (ORF405).  
 GN CYP109.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M23;  
 RX MEDLINE=91192601; PubMed=1849493;  
 RA Ann K.S., Wake R.G.;  
 RT "Variations and coding features of the sequence spanning the  
 RT replication terminus of Bacillus subtilis 168 and M23 chromosomes";  
 Gene 98:107-112(1991).  
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED  
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M24523; AAA22720.1; -.  
 DR HSSP; O5080; 1107.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KM Oxidoreductase; Monooxygenase; Electron transport; Heme.  
 FT METAL 351 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 405 AA; 45845 MW; 1331D5BEA74E3C04 CRC64;  
 Query Match 19.8%; Score 420; DB 1; Length 405;  
 Best Local Similarity 29.0%; Pred. No. 2.5e-19;  
 Matches 121; Conservative 72; Mismatches 182; Indels 42; Gaps 9;  
 QY 1 MFOEANOSETKPAFPKFPAGYAD---PPAIERERREATPIFYWDEGRSWLTRYH 56  
 DB 1 MTNQTARSSKKEKYANLIMEELHSKDLFPPIYDKLRRESPVYDRLRCMDVFXKD 60  
 QY 57 DVSATFRDERFAVREWESSAEYSAPLELSDMKKYGFLPGLEPDHARVKLVNPSFSS 116  
 DB 61 DVQFVLKNPKLFFSKRGIQTES-----ILTMDPPGHTLRALVSPAFRP 104  
 QY 117 RAIDLRAEIQRVVDQLDARSQGEFVVRDYAEGIPKRAISALLKVPACDEKFRFG 176  
 DB 105 KAVKQLETRIKNQVTAFLQEARQKSTIIDIEFAGPLVITIAEMGAIERHLLIKTYS 164  
 QY 177 SATARALGVGLVQVDEETKTLVASVTEGLALP---HDVLDERRRNPLENDVLTMLQ 232  
 DB 165 D---VLVGAKDSSKAVADWVHNRDGHAFSLDFRILSKRAEPKE-DLMTMLQ 219  
 QY 233 EADGSRSLSTKELVALVGAIIAAGTDTTITVLIAPAVLNLSPEALELVKAEPLGRNALD 292  
 DB 220 EIDGEVLTBEQILGFCILLVAGNETTTLNLIANAVALYLTEDSVQOVQONTDNVANVIE 279  
 QY 293 EVLR-FDNILRIGCTVFARODLEYGASIKKGEWVFLIPSLRDOCTVPSRPDVDR 351  
 DB 280 ETLRYVSPVQAIG--RVATEDTELGVFTKKSSVSIWASANRBDKCKCDKCFKIDRP 337  
 QY 352 TGASLAYGRGPHVCPGVSRLAEIAVGTIF-----RRPEMKLKETP---VFG 398  
 DB 338 SYPHLSFGFGLHFCGLAPLARLEANIALSSLSMSACIKRAHDELTLEAIPSFVVG 394





DE Cytochrome P450 55A3 (EC 1.14.-.-) (Cytochrome P450NR2).  
 GN CYP55A3.  
 OS Cyllindrocarpum konkinense (Cyllindrocarpum lichenicola).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; microsporite Hypocreales; Fusarium.  
 OX NCBI\_TaxId=42744;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO 30561;  
 RX MEDLINE=97163854; PubMed=9010609;  
 RA Kudo T., Tomura D., Liu D.L., Dai X.Q., Shoun H.;  
 RT "Two isozymes of P450nor of Cyllindrocarpum konkinense: molecular  
 RT cloning of the cDNAs and genes, expressions in the yeast, and the  
 RT putative NAD(P)H-binding site";  
 RL Biochimie 78:792-799 (1996).  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D78512; BAAL1409.1; -  
 DR PIR; JCS674; JCS674.  
 DR PIR; PD0007; PD0007.  
 DR HSSP; P23295; 1F24.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Oxidoreductase; Monooxygenase; Heme.  
 FT METAL 357 357 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 SQ SEQUENCE 408 AA; 45381 MW; E99298B1949E7E5 CRC64; (BY SIMILARITY).

Query Match 18.2%; Score 386.5; DB 1; Length 408;  
 Best Local Similarity 28.6%; Pred. No. 3.1e-17;  
 Matches 116; Conservative 68; Mismatches 174; Indels 47; Gaps 9;

QY 7 NSETPAFDFKPPAGVADPPPAERLRERATPPIF---YWDGKRWVLTIRHDSAVFR 63  
 DB 6 DETTTPRF---PFORASAFEPAPPARLRANEPISQVELFDQSLMLVVKHEDVCRAVAT 62  
 QY 64 DERFAVSREWESSAAYSAIPELSPMKYVGLGLP-----PEDARVRKLVNPSFTR 117  
 DB 63 DERLSKERR-----LGFPELSGGAANKKFTFVDMADPAMNORSWEPEFTED 114  
 QY 118 AIDLRAELIQTVDOLD--ARSGEEDVVDVAEGIPMRASALIKVPAECDE--- 170  
 DB 115 HVENLRPYIKETVQGLNDMVANGCEPDLIEKFAIPVPSYITITLGVPPEDLEYLNE 174  
 QY 171 --KFRRFSGATRALGVGLVPOVDETKVLVASVTEGALLHDVDERRRNPENDVLTW 228  
 DB 175 QNAIRKNGSGTA-----CEAAAMQCLIKYLAKLVQRIQEP-KDILLIR 218  
 QY 229 LFOAEADGSRSLSTKELVALGAILAGDTTITLFAVNLILRSSEALEIKVAGKMR 288  
 DB 219 LVDOQLVPHIEKSDVVOIAFLVAGNANMMIALGVTTLMQNSQLEELKADPTLP 278  
 QY 289 NALDEVLRPNILRTGVTFARODLEYCGASIKKGMVFLILPSALRDGVTSRPDPV 348  
 DB 279 GVEVELCRVHTGSSAMKRVAKEDMELGKILRAGSGIIASNSGANDRDVFPNPVFM 338  
 QY 349 KRDGTGA--SLAYGRGHVPCVSLALELEIAVGTIFRRPPEMKL 391  
 DB 339 HRDFPSRDGLGFGFPHRCIAELIAKAELEIVETLFAVLPLDURV 383

DT 30-MAY-2000 (Rel. 39; Created)  
 DT 30-MAY-2000 (Rel. 39; Last sequence update)  
 DT 28-FEB-2003 (Rel. 41; Last annotation update)  
 DE Putative cytochrome P450 142 (EC 1.14.-.-)  
 GN CYP142 OR RV3518C OR MT3619 OR MTW023.25C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxId=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634220;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulton J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence";  
 RL Nature 393:537-544 (1998).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleisemann R.D., Alland D., Bisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL022022; CAAL1775.1; -  
 DR EMBL; AB007165; AAK4797.1; ALT\_INIT.  
 DR PIR; H70807; H70807.  
 DR HSSP; P33006; 1CPT.  
 DR TIGR; MT3619; -  
 DR TubercuList; RV3518C; -  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;  
 KW Complete proteome.  
 FT METAL 340 340 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
 FT CONFLICT 1 18 MTEADVDLADGNFYASR -> GWSPNRSRAPPSRAPRP  
 FT (IN REF. 2).  
 SQ SEQUENCE 398 AA; 44398 MW; BCF3C23CEB5767F CRC64;

Query Match 18.2%; Score 386; DB 1; Length 398;  
 Best Local Similarity 29.3%; Pred. No. 3.3e-17;  
 Matches 119; Conservative 65; Mismatches 168; Indels 54; Gaps 11;

QY 36 REATPIFYWDGSRVLTIRHDSAVFRDFAVSREWESSAAYSAI-----PELSDM 90  
 DB 18 REAPAAVYMRAN-----QVPRFD-----RNLGAASATQVAVDAERQPELS- 60  
 QY 91 KKYGLFGLPPED-----HARVKLVNPSFTSRAIDLRAELIQTVDOLDAR 137  
 DB 61 --NAGGIRPDOPALPMIMIDMDPAHLRRKLVNAGFTKRVKDXEASIALCDTLIDAV 117





```

Db 61 -----SKVTRPGPELNLNAGGKQAKAPTVDNDADPHNMQRGVESLFLTEHYKIQ 114
Qy 124 AEIQRTVDQLDARS-----GQEFVVDVDAEGIPMRRAISALLKVP-----AECDEKPR 173
Db 115 PYIQTVDDLLAMKKKGCANPVDLVKGFALPVPSYIITLIGVPFNDLHLTQNMNR 174
Qy 174 RGSATARLGLVLPQVDEEFTKLVASSTEGALLHDVLDERRRNPENLVLTMLQAE 233
Db 175 TNGSSTARAS-----AANQELLDYLAISVERLEBP-KDDLSIKLCTEQ 218
Qy 224 ADGSLSTELVALVGAIIAGTDTIYLAFAVNLNLSPEALVELVKAPEGLMNADE 293
Db 219 VKPGNIEKADAVQIAFLVLVGNATVNMNIALGVVTLFQHPQLAQKPNPSLADQFVEE 278
Qy 224 VLRFDNLRIGTVRPARODLEYCGASIKKGEVFLIPSLRDGTVFSRPDVFDRPDGT 353
Db 279 LCRVHTASALAIKFRANKVDLEIGGHIKANEGLIANSQANDADLIFENPDEFNNRKP 338
Qy 354 AS-LAYGRPHVPCGVSLARLEAIVNGTIFRRPEM-----KLKTPVFG 398
Db 339 AEDPLGYGFPRIAEHLAKAELTTFVATLFKEFPDLNIAVPFEKINFTPLGG 392

```

```

RESULT 15
C128_MYCTU STANDARD; PRT; 489 AA.
AC Q59572;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 128 (EC 1.14.-.-).
GN CYP128 OR RV2268C OR MT2330 OR MTCY339.42.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekala F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; 277163; CAB00967.1; -.
DR EMBL; AB007076; AAK46612.1; -.
DR PIR; H70729; H70729.
DR HSSP; 000441; 10XA.
DR TIGR; MT2330; -.
DR Tuberculin; RV2268C; -.
DR InterPro; IPR01128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Transmembrane; Complete proteome.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 315 335 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT METRL 435 435 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ
SEQUENCE 489 AA; 5313 MW; 633F233CEFD3AD7A CRC64;

```

```

Query Match 17.6%; Score 373.5; DB 1; Length 489;
Best Local Similarity 28.3%; Pred. No. 2.5e-16;
Matches 115; Conservative 65; Mismatches 204; Indels 23; Gaps 6;

```

```

Qy 16 DFKPPAGYADPPPAIRLEAPPIFYWDGSRVLTTRYHDVSAVFERDPAVSREWE 75
Db 101 DFDPPDAIADPPYHYRELLAGERVQYPRDVIYLSRYADVRAAHNDTLSSARGVT 160
Qy 76 SSAEYSSAIPPLSDMKKGLGFLPREDPAVRKLVNPSFTSRALDILAELQRTVDQLD 135
Db 161 FSRGMLPFLP-TSD-----PPATTRRKQIAPGMARGALETPMVDQLARELVG 209
Qy 136 ARSGQEEVDVVDYAEIGIPMRRAISALLKVPACDEKFFRFGSATARALGVLPQVDEET 195
Db 210 GLUQTTPADVSTYAAHPMRRAISVGLVDGPDDEAFRLSNQARITDVAL-----S 262
Qy 196 KTLVASVTEGLA---LHDVLDERRRNPENLVLTMLQAEADGSRISTELVALVGA 251
Db 263 ASGLISTVQGFAGFRRLALFTHRDNGLRECTVIGKLTATHAEGRLSDDELFFFAVL 322
Qy 252 IAAGDTTIYLAFAVNLNLSPEALVELVKAPEGLMNADEVRFDNLRIGTVRPARQ 311
Db 323 LVAGESTAHMISTFLTLADYPPQLTLAQOQDILPSAIEHNLRFISPIQ-NICRTTRV 381
Qy 312 DLEYCGASIRKGEVFLIPSLRDGTVFSRPDVFDRPDGTGASLAYSRGPHVPCGVSLA 371
Db 382 DYSVQAVIPAGSLVLLAMGAAANDPROYEDPVRADARNVGHILAFSGSGLHLCRTQLA 441
Qy 372 RLEAEIAVGTIFRRPEMKLKETPVFGYHAPAFRNIESLVILKPSKA 418
Db 442 RMEGQALIRELVANIDRIEIVEPPTWTJNANLRLGLTRVAVTPRVA 488

```

```

Search completed: October 2, 2003, 17:28:42
Job time : 9.74239 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Gen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:33 ; Search time 14.5739 seconds  
(without alignments)  
2764.850 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124

Sequence: 1 MTOEQANQSEKPAFDFKPF.....HPAFNIESLNLKPSKAG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	531	25.0	405	1 B42606	cytochrome P450 CV
2	525.5	24.7	395	1 G68594	cytochrome P450 B1
3	518.5	24.4	396	1 B63851	cytochrome P450 YJ
4	502	22.6	410	1 E69611	cytochrome P450 CY
5	486	22.9	376	1 G69679	polyketide hydroxy
6	481.5	22.7	418	2 G83229	cytochrome P450 PA
7	459	21.6	403	2 B35401	cytochrome P450 10
8	451	21.2	415	2 AC2731	cytochrome P450 HY
9	451	21.2	415	2 B97512	probable cytochrome
10	450	21.2	402	1 A70707	cytochrome P450 RV
11	442.5	20.8	438	2 C87170	probable cytochrome
12	434.5	20.5	434	1 E70515	cytochrome P450 RV
13	430	20.2	411	1 J54003	cytochrome P450 -
14	429	20.2	417	1 S49051	cytochrome P450 CY
15	429	20.2	411	2 T36526	probable cytochrome
16	421.5	19.8	444	2 C83336	probable cytochrome
17	420	19.8	411	1 J55859	polyketide synthase
18	415.5	19.6	402	2 H82813	cytochrome P450-11
19	413	19.4	397	1 S51594	cytochrome P450 MY
20	411	19.4	404	2 T30231	cytochrome P450
21	406	19.1	310	2 T44767	cytochrome P450 (i
22	400	18.8	406	1 S18531	cytochrome P450 ER
23	386.5	18.2	408	2 PD0007	cytochrome P450 NO
24	386.5	18.2	408	2 J55859	cytochrome P450 NO
25	386	18.2	398	1 H70807	cytochrome P450 RV
26	380	17.9	381	1 S15809	cytochrome P450 CY
27	373.5	17.6	489	1 H70729	cytochrome P450 RV
28	366	17.2	410	2 S3924	cytochrome P450me9
29	364.5	17.2	399	1 A55578	cytochrome P450 -

30	361.5	17.0	412	2 S24750	cytochrome P450 (a
31	361.5	17.0	413	2 T44587	cytochrome P450 HO
32	357.5	16.8	430	2 J54287	cytochrome P450 BC
33	356.5	16.8	400	2 B82817	cytochrome P-450 h
34	354.5	16.7	406	2 A35401	cytochrome P450 10
35	353.5	16.6	402	2 J55150	nitric-oxide reduct
36	352.5	16.6	404	1 J55150	nitric-oxide reduct
37	352.5	16.6	405	1 J55150	cytochrome P450 CY
38	350	16.5	401	1 H70921	cytochrome P450 BJ
39	344.5	16.2	400	1 H70921	cytochrome P450 RV
40	344.5	16.2	433	1 B70677	cytochrome P450 RV
41	342.5	16.1	428	1 A42971	cytochrome P450 ER
42	338.5	15.9	444	2 F87366	cytochrome P450 FA
43	332	15.6	410	1 O4BS6M	cytochrome P450 10
44	328.5	15.5	406	1 A48495	linalool 8-monooxy
45	328.5	15.5	467	2 T10879	y41c protein - Rhi

## ALIGNMENTS

RESULT 1  
B42606  
cytochrome P450 CYP11B1 - Saccharopolyspora erythraea  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Saccharopolyspora erythraea  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Mar-2000  
A:Accession: B42606  
R:Andersen, J.F.; Hutchinson, C.R.  
J. Bacteriol. 174, 725-735, 1992  
A:Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and enzy  
A:Reference number: A42606; MUID:92121109; PMID:1732208  
A:Accession: B42606  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA; protein  
A:Residues: 1-405 <AND>  
A:Cross-references: GB:M83110; NID:G152682; PID:AAA26483.1; PID:G152684  
A:Experimental source: NRRL2338  
A>Note: sequence extracted from NCBI backbone (NCBI:77484)  
C:Genetics:  
A:Gene: CYP107B1  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: Chromoprotein; heme; iron; metalloprotein; oxidoreductase  
F:328-374/Domain: cytochrome P450 homology <CYP>  
F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted  
Query Match 25.0%; Score 531; DB 1; Length 405;  
Best Local Similarity 32.5%; Pred. No. 6.8e-30;  
Matches 132; Conservative 83; Mismatches 169; Indels 22; Gaps 6;  
QY 16 DEKPAFGYARDPPAIRLRRE--ATPFYDEGSMVLTGYHDVSAFRRDER----RAV 69  
8 DLAFDPAFADDRNNRVARKEEPVQRTYVNGLDAMITREYDQALLPRLAKDGR 67  
QY 70 SREWESSAEYSAPISLDMKKYGLFGLPPEDHARVRLVNPSTSAIDILRAEIORT 129  
68 TQOIERKLADARRRPGSPDLGPMMLNTDPPDHRLAKLVKATARVEELRRIRQI 127  
DB 130 VDOLLARSQGEEDVDVDAEGIPMRALISALLKVPACDEKFRFGSATARALGVGLVP 189  
128 TDLDLLRLRGLSEVDLIDFAFPLPIYISLGMGVEDSRDDFRSM---TNVLVDGSP 183  
QY 190 QVDEERTK-TLVASVTEGALLHDVDERRRRPLENDVLTMLQAEADSRISTEKVLV 248  
184 EAQAQASVAMVEYLTETLA-----KKRTPEGDDLTALLEAVEDDRLRSEGLIMV 235  
QY 249 GAIAGTDTTIVYLAFAVLNLRSPALIEVKAEPGLMRNALDEVLFNDRILRIGTVRF 308  
236 FLVLVAGHETTVNLIGNCVSLGNPDQALALRNDPSLPPALIEETTLAYESPVANGTFRH 295  
DB 309 ARDLEYGASIKKGEVFLIPSAALRDGVFSRRPDVVDVRRDGSALAYGRPVCGV 368  
296 TAAAVRFGDVVPIPGELVWVALGAAANRDEEFEDPDRDITRETTGHAFAFGHIFCVGA 355

QY 369 SLARLEAIVGTIFRRPFEMKLKETP--VGYHPAFRNIESLNV 411  
:|||||:| : ||||: : : |  
Db 356 ALARLEAQIAVGRLLERFDLRMAASPDDLRRFRFSVLMRGLEKLPV 401

## RESULT 2

cytochrome P450 biol - *Bacillus subtilis*  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: *Bacillus subtilis*  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: G69594

A.Authors: Fouglier, D.; Fritz, C.; Hilbert, M.; Fujita, Y.; Puma, S.; Gallizzi, A.; Gallelli, I.; Harwood, C.R.; Henaux, A.; Hildebrandt, H.; Holsteppel, S.; Hosono, S.; Hullio, M.F.; Koetterer, P.; Koningsstein, G.; Krogh, S.; Kumano, W.; Kurita, K.; Lapidus, A.; Lathionis, R.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauelet, Y., M.; Ogawa, K.; Ogiwara, A.; Oudgaa, F.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelecchi, Riger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serovani, A.; Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serovani, A.; kench, M.; Tamakoshi, A.; Tanaka, T.; Terpest, P.; Togmoni, A.; Tosato, V.; Uchiyama, T.; Winfers, P.; Wipal, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, A.; Authors: Yoshioka, H.F.; Zunshtein, E.; Yoshiokawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A|Reference number: 69580; MUID: 98044033; PMID: 9384377

A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-395 <KUN>  
A:Cross-references: GB:Z99119, GB:AL009126, NID:G2635411, PIDN:CAB14997.1, PID:G2635503  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: bio1  
C:Superfamily: Bacillus cytochrome P450 CYP106, cytochrome P450 homology  
C:Keywords: oxidoreductase  
E:123-367/Domain: cytochrome P450 homology <CYP>  
E:123-367/Domain: cytochrome P450 homology

Query Match	24.7%	Score 525.5	DB 1	Length 395
Best Local Similarity	31.9%	Pred. NO. 1.6e-29		
Matches 129	Conservative 98	Mismatches 137	Indels 41	Gaps 12

```

QY      24  YAEDEPPAIEKREATPIFYMBGR-----SWLTRHYDVAVFPRDERFAVSREEMSSA  78
      :  ::  ||  ::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      12  FLKNPYSFYDLRAVHPYI--KGSFLKYPGWYVTGYEETAAALKDARFKRTPPLESST  68

```

```
OY      79 EYSASIPELSDMKKYGLFGLEPPEDNARVRLKLVNSFTSRALDILRAEQRTVQDLDRS 138  
       :|::: |||| ::| ||  
Db     69 KYQ----DSHVQNQMMLFQNQRDHNRRLTASGAFFRPTTESYQRYIETVAHNLDDVO 124
```

```

QY      139  GQEEEDVVRVYAEGLPMRAISLTKVPAECDEKFRRGSAATARALGVGLVPQVDEETKL 198
        |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db      125  GKKEEVIISPAFPLASFLVIANIIGVPEEDRQLKEMAAS-----LIQTID-FTFSR 175

```

```

QY 199 VASATGEG-----LALLHVDLDERRRNPLENDVLTMLLQAEDSGSRSTKELVALVGA I 251
      | : ||| : :: : | : | : :: ||| : : : | : :
Db 176 KA-LTEGNIWAQAAYFKELIQKRKHP-QQDMISMILKGR-EKDKYTEEAASCTILL 232

```

Dy 252 IAAGDITTYLIAFAVLNLRSPAEALVKAPGLMKNALDELVRFDNIIRIGTVRFARQ 3111

Dd 233 AIAGHTTVNLISNSVLCIQHPQLKLRENPDILGVAEECLRYESPOTW-TARVASR 2911

[illegible]

```
QY      372 RLEAEIAGVTIFRRFPPEMKLKE-----TPVFGYHPAFRNIESLN 411
           |||::: | : | : | : |||   | : |
Db      352 RLEAQIAINTLTQRMPSLNLADFEWRYRPLEG----FRALIELPV 392
```

```

RESULT 3
Be9851
cytochrome P450 yjIB - Bacillus subtilis
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: B69851

```

A.; Ehr1  
Nathu 2

Nature 350, 297-298 (1993).

A:Authors: Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Funo, S., Galizzi, A., Gallardo, J., Harwood, C.R., Hentut, A., Hilbert, H., Holtsappel, S., Hosono, S., Hultio, M.F., Iach, U., Königstein, G., Krogh, S., Kumano, M., Kunita, K., Lapidus, A., Lardiosso, P., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Maeda, S., Manasek, Y. M., Ogawa, K., Oglavara, A., Oudasa, B., Park, S.H., Parro, V., Polh, T.M., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T., Scanlon, J., Sekowska, A., Sericova, S., Schroeter, R., Scoffone, F., Seikuchi, J., Yokosaka, A.:Authors: Schlecht, S., Schwoerer, R., Scoffione, F., Seikuchi, J., Tosato, V., Uchiyama, T., Winere, P., Wipat, A., Yamamoto, H., Yamane, K., Yasunoto, K., Yata, K., Yoshida, A.:Authors: Yoshikawa, H.F., Zumbel, E., Yoshikawa, H., Danchin, A.

Allite: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
Reference number:A69580; MUID:98044033; PMID:3384377

A:Accession: B69851  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-396 <KUN>  
A:Cross-references: GB:Z29110; GB:AL009126; NID:z6533472; PIDN:CAB13078.1; PID:z6533575  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yjiB  
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
F:356-371/Domain: cytochrome P450 homology <CYP>

Query Match	24.4%	Score 518.5	DB 1	Length 396
Best Local Similarity	32.6%	Pred. No. 5e-29		
Matches 124, Conservative	72	Mismatches 155	Indels 29	Gaps 6

QY 25 AEDDEPFAIRLRREATPIFYWDEEGSWLTRYHDVSAVFDERFAVRSREEMESSAIESSAI 84  
Db 23 AYHDFPWYESMRKDAFVSFDEENQVSVLLYDDVKKVGVGDEL-----FSSCM 70

Qy 85 PELSDMKKYGFGLEPEDARVRKLVNPSFTSRALDLRAEIQRYVDLIDARSSQEEFD 144  
71 PQTSTSGNSIINNDPKTKTKIRSVNKAFTPRVMQKMERIQETDDELQKQFSRSEFD 130  
Db

Oy 14 VVRPYAEGIPMRATISALILKPAEDCEKFERFGSATARALGVGLVQVDEETKTLVA---- 200  
 131 LVHDFSYLPPIVISELGLVPSAIMEQFKAMSD-----LLVSTPKDSEAEKAFLEERD 185  
 Db

Oy 20 SVETGIALTHDVLDERRRNP LENVLTMLQAEKDGSRSTKEI VALVGA IIAAGDTTI 260  
Db 186 KCEELAAFFAGRIIEKRNRKPEQIIISILVEAEEGEEKSGEELLPFCTILLVAQNETT 245

Qy 261 YLILFAVNLRLRSPDALFLVKAEGLMRNALDVEVRF-----DNIRIGTGFARADLEYCG 317  
Db 246 NLISNAMYIILETGVEELRSHEHLMPOAVEALRRFAPAVLR-----RIAKRDTEICG 301

QY ASIKKGEAMFLIPSAIRDGVESRPDVEVRARPDGASLAVGRGHVCPGSVLARLEAEI 377  
||| : ||| :  
318 | : : : : :  
Db HLIIEGDWNLAFVASANRDEAKFDRPHMEDIRHNHNPHTAIFGHGIHFCLGLAPLARLEANI 361

Qy	378	AVGTIFRRPEPMK-LKETPV	396
		: : :     :	
Db	362	ALTSLSIAEPHMECVSTPI	381

RESULT 4  
E69611  
cytochrome P450 cypA - *Bacillus subtilis*

## RESULT 4

E69611  
cytochrome P450 cypA - *Bacillus subtilis*  
N;Contains: oxidoreductase (EC 1.-.-.-)

[illegible]

RESULT 5  
669679  
polyketide hydroxylase (EC 1.-.-.-) pksS - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: 669679  
A:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertecchi, C.; Bron, S.; Brönnle, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chai, E.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Poulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gallardo, J.; Harwood, C.R.; Hénaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hulo, M.F.; Koetter, P.; Königstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lander, U.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Mauee, Y. M.; Ogawa, K.; Ogizawa, A.; Oudsga, B.; Park, S.H.; Parro, A.; Pohl, T.M.; Portetelli, R.; Rieger, R.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadle, Y.; Sato, T.; Scanlon, T.; Schleich, S.; Schroeder, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Authors: Schleich, S.; Schroeder, R.; Scoffone, P.; Togouchi, A.; Tosato, V.; Uchiyama, A.; Winberg, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, A.; Authors: Yoshikawa, H. F.; Zunnstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: 669679  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-376 <KUN>  
A:Cross-references: GB:Z59113; GB:AL009126; NID:G26340930; PIDN:CAM13607.1; PID:G2634107  
A:Experimental source: strain 168  
C:Genetic: pksS  
C:Gene: pksS  
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:Keywords: oxidoreductase  
C:40-376/Domain: cytochrome P450 homology <CYP>

Query Match	22.9%	Score 486	DB 1	Length 376
Best Local Similarity	31.6%	Pred. No. 9,28-27		
Matches 120	Conservative 76	Mismatches 156	Indels 28	Gaps 7
Qy	17	FKPAPGVAEDPPPAIRLRERATPIFYWDEGR-----SWVLTRYHDVSAVFRDERFA--	68	
Db	8	FHPGKEFHNNPFPEVLGRFREEEPIDHRFELKRFQATYPAWVITRYDDCMALDKDNRITRD	67	
Qy	69	---VSREWESSAAYSAIPELSPMKRYGGLGEPREPHARVRKVVPNSFTSRATIDLRAE	125	
Db	68	VQVVMNNOEQIKMLNVSIEDIDFVSD---HMLAKTTPDHTRLSLVHQAFTTRITENLNGS	123	
Qy	126	IQRTVDLLDLARSQGEFEVDVVDVAEGIPMKRAISALKLVPAECCEKFRFRFGSATARAL--	183	
Db	124	IEQIAEQDLDMEMKENKADINKSPASPLPFIIVISELNGIPREDSRQFIQWVNAVVDTEG	183	
Qy	184	GVGLVPOVDDETKTLVLSVTEGLALLHDVLDERRRNPENDVLTMLQAEADGSRLSXKE	243	
Db	184	NRELTNQALRFKKYIAK-----LIHD---RKIKP-KODLLIKLVAAEENGSKLSKE	232	
Qy	244	LVALVGAIAGTDTTYLTALFAVNLTLRSPEALELVKABEGLKRNALDEVLRGPNITRI	303	
Db	233	LYSMLFLVVAAGLETTYNLGGSTLALLQHKKEBKLRKQCEMTATAVEELRTTSPPVM	292	
Qy	304	GVYFARQDLEYCGASIKKGEVFLIIPSLARDGTVPSPDVPDRDGTASLAYGRGP	363	
Db	293	MANMWAIEDFYKSHSIRKGDMLFIFIGISANRDNPENPEPILININSPPNHISFGFCIH	352	
Qy	364	VCPGVSLARLEAETIAYGTIF	383	
Db	353	FCUGAPLARLEGHATPKAIF	372	

RESULT 6

G83229

Cytochrome P450 PA3311 (imported) - Pseudomonas aeruginosa (strain PAO1)

CiSpecies: Pseudomonas aeruginosa

CiDate: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 16-Feb-2001

C/Accession: G83229  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
N:ature 406, 959-964, 2000  
A:/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
A:/Reference number: A82950, MUID:20437337; PMID:10984043  
A/Accession: G83229  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1,418 <STO>  
A:/Cross-references: GB:AE004755; GB:AE004091; NID:g9949456; PIDN:AA06719.1, GSPDB:GN001  
A:/Experimental source: strain PA01  
C/Genetics:  
A:/Gene: PA3331  
C:/Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
P:/333-369/Domain: cytochrome P450 homology <P45>

Query Match 22.7%; Score 481.5; DB 2; Length 418;  
Best Local Similarity 33.5%; Pred. No. 2,2e-26;  
Matches 116; Conservative 73; Mismatches 140; Indels 17; Gaps 5;  
QY 50 WVLTRHVSAPVDERFPAVSREBESSAAY---SAIFELSDMKKYGFLGPPEDHAR 105  
DB 41 WVTTRYDARKVNLNHP--GVRDARQAALYAKRTGSPRAGIGSGLSHMLNLDPPDHTR 98  
QY 106 VRLVNPSTFRAIDLRARIORTVDQLARSGOEFDPVVRVYAGGIPMRATISALIKVP 165  
DB 99 LRLVGRAPFPPOVERQPIHETITTEALDAMGRBQADLADPAIFLTAVIFELGIP 158  
QY 166 -AECDEKRRFSGATARALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDV 225  
DB 159 EAREHARQSGWQAEELS-----PEAQALADADVLDLRLV---LEAKRQPAD-DV 207  
QY 226 LTMLOAEADGSLSTELVALYGAITAACTDTTIIIAAVNLRSPPALBELVAKBPG 285  
DB 208 YSGIVQAADSGQLSEAEIVSMHMLMSGFEETMMIGALVTLLVNPQALLRQPE 267  
QY 286 LMGNALDEVLPFNILRIGVRFAPRODLECGASIKKGEWVFLIPSLDGTVFSPDV 345  
DB 268 LTNAAEELVRHDSPVASMLRFTVEDVLDGVTIPAGEYILVSNLTANHDAERFDDPR 327  
QY 346 FVVRDGTGASLAVGRPHVCPVSLARLEAIVAGTIFRRPEMKL 391  
DB 328 LDTLRNTDGLTGFGVHYCVGASLARLEGRIMQRLARFPDQL 373

RESULT 7  
B35401  
cytochrome P450 105B1 - *Streptomyces griseolus*  
N:Contains: oxidoreductase (EC 1.1.1.17)  
C/Species: *Streptomyces griseolus*  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 04-Mar-2000  
C/Accession: B35401  
R:Omer, C.A.; Lénstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romesser,  
J.; Bacteriol. 172, 3335-3345, 1990  
A:/Title: Genes for two herbicide-inducible cytochromes P-450 from *Streptomyces griseolus*  
A:/Reference number: A35401; MUID:50264332; PMID:2345149  
A/Accession: B35401  
A:/Residues: 1-403 <OME>  
A:/Molecule type: DNA  
A:/Cross-references: GB:M32239; NID:g153480; PIDN:AAA26825.1; PID:g153481; GB:M36481  
C/Genetics:  
A:/Gene: CYP105B1  
C:/Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;  
F:/238-374/Domain: cytochrome P450 homology <P45>  
F:/352/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.6%; Score 459; DB 2; Length 403;  
Best Local Similarity 31.7%; Pred. No. 8.1e-25;  
Matches 127; Conservative 73; Mismatches 157; Indels 44; Gaps 12;

QY 21 ARGVADPEPPAIERLREATPI---FYWDEGRSWLTRYHVSAPVDERFPAVSREBESS 77  
DB 20 AFGCPDPAPADVTEAARTPTVRATLMDGSSCMLVTRHODVRAVLGDPPF-----SA 71  
QY 78 AERYSAIPELSDMKKYGLFGLP-----EDHARVKLVNSFTRALDILRAEIORV 130  
DB 72 DMRGTFFPLTIGGR-EIIGTNPFLRMDPPHARRRRLTADFTVKYKAEAKPEVQRLA 130  
QY 131 DOLL-D-ARSGOEFDPVVRVYAGGIPMRATISALIKYPAECDEKRRFSGATARALGVGLV 189  
DB 131 DDLVDMMTKGRSADLVTEFALPLPSVLICLLGVPYE-DHAF--FQESRVLLTLRSP 187  
QY 190 QVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLLOEADGSLSTELVALYG 249  
DB 188 ---EEVRAAQDELEIYARL---ARTKRPRDDAISRVLARGELDDTOIATGRLTL- 240  
QY 250 AITAACTDTTIIIAAVNLRSPPALBELVAKBPGMKNALDEVLPFNILRIGVRA 309  
DB 241 ---AGHETTAMNTALSTLVLRNPDQLARLAEPAVLGAABELRYTLTVHNGVRLA 296  
QY 310 RODLEYCGASIKKGEWVFLIPSLARDGTVFSPRPVDFVRDPTGASLAVGRPHVCPGV 369  
DB 297 TEDVLIGRTIAGSEVLCMISSANRDAVFPFGGDDLDVARDARRVARFGVHQLGCP 356  
QY 370 LARLEAIVAGTIFRRPEMKL---KETP-----VFQYH 400  
DB 357 LARVELQIAIETLRLRLPLRLAVPHEIIPFGMDMAYGVH 397

RESULT 8  
AC2731  
cytochrome P450 hydroxylase Atu1256 [imported] - *Agrobacterium tumefaciens* (strain C58,  
C/Species: *Agrobacterium tumefaciens*  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C/Accession: AC2731  
R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavain, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A:/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AC2731  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-415 <KUR>  
A:/Cross-references: GB:AE008668; PIDN:AAU42265.1; PID:g17739662; GSPDB:GN00186  
A:/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A:/Gene: Atu1256  
A:/Map position: circular chromosome  
C:/Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology  
C:/Keywords: heme; iron; metalloprotein  
F:/367/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.2%; Score 451; DB 2; Length 415;  
Best Local Similarity 29.1%; Pred. No. 3.1e-24;  
Matches 118; Conservative 82; Mismatches 175; Indels 30; Gaps 9;

QY 22 PGVADPPFAIERLEATPIFYWDEGRSWLTRYHVSAPVDERF-----AVSREB 75  
DB 24 PARVNDPFPVYALAHQCFYWEQORWFCTCYDHVSTLLDRRGRGOILHVASREE-- 81  
QY 76 SSAEYSSAIPF-LSDMK-----KYLFGFLPDPDARVKLVNPSTFTRALDILRAEIOR 128  
DB 82 -----IGPPELEHVKHFDLAEQHSLLLEPEHTRTLTLNRAFSVSHVDMKPEIEB 135  
QY 129 TVDQLDARSGOEFDPVVRVYAGGIPMRATISALIKYPAECDEKRRFSGATARALGVGLV 188  
DB 136 LANRLAEAPANGCTELISSYADIIPYMIARITGIPEDMGOLKMSHA---YGVWYMF 192  
QY 189 PQVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLLOEADGSLSTELVALV 248



D<sub>b</sub> 193 KRTEDDELLAKKAOEFSDDYVRVYAERRAP-KDOLLSHMITHHKQIYTDELVESTT 251  
Q<sub>y</sub> 249 GAIIACTDTITYLFAFANILNLRSPALBLVKAEFGMLRNALDEVLFNDILIRIGTVRF 308  
D<sub>b</sub> 252 IVLNAGHEATVHQGSVRRIIBESGDPKTLFHDETERTVEETLLICAPWH-PRW 310  
Q<sub>y</sub> 309 ARODLVCASIIKKGEWFLLIPSAIRDGTAFSRPDEVDRDITCASLAYGRGPVPCGV 368  
D<sub>b</sub> 311 VLEPEITGVGFQRDKSLILMAANDPAKFSDPLAQPPRNBGANVSFGAIGHFCIGA 370  
Q<sub>y</sub> 369 SLARLEAIVGTFIFRRPEMKLKETPVFG--YHFAFNIESLV 411  
D<sub>b</sub> 371 PLARIELNALPLFKRLPGLIKAEPPVKOVTH-FGLERLDL 413

RESULT 9  
B97512  
Probable cytochrome P450 hydroxylase [imported] - Agrobacterium tumefaciens (strain C58)  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
C/Accession: B97512  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Guirillo, B.; Goldman, A.; Liu, F.; Mollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Matkeiz, B.  
Science 294, 2223-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A/Reference number: A97359; MUID:2160851; PMID:11743194  
A/Accession: B97512  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-415 <KUP>  
A/Cross-references: GB:AE007869; PIDN:AAK87051.1; PID:q15156303; GSPDB:GN00169  
C/Genetics:  
A/Gene: AGR\_C\_2319  
A/Map position: circular chromosome  
C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homolog  
C/Keywords: heme; iron; metalloprotein  
F/367/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

Query Match      21.2%: Score 451; DB 2; Length 415;
Best Local Similarity 29.1%: Pred. No.3,1e-24;
Matches 118; Conservative 82; Mismatches 175; Indels 30; Gaps 9;

QY      22  PGVADPPPAERLREATPIFYWDEGRSWLTRYHDAVAFDERF-----AVSREBWE 75
Db      24  PAFYNDPWPVYAALHAQCPTFYWEEQRMFFCYDHVSTLLRDRRFRGQILHVASRE-- 81

QY      76  SSAEYSSAIFR-LSDMK-----KYGLFGLRPEDHARARKLVNPSFTRAIDLRAEIQR 128
Db      82  -----IGLPRLPLeVKhFDLAEOHSLLEBPETHRTLRITINRAFVSRHVDKMKPELAE 135

QY      129  TVDOLLDRASQOEEDVDVRYDAEGIPMRATISALTIVPAECDEKFAFPSSATARALAGVGLV 188
Db      136  LANLILIEFANGETRELLISSYADILPVTIMARMIGIPEEMGRQLLKMSHA---YVGWTFM 192

QY      189  PQVDEETKTLVASTEGTGLALHDVLDERRRNPLENDVLTMLIQAEADSGRLSTKELVALV 248
Db      193  KRTEDEDLADKAQEFSDYVRVRYAERRAEF-KODLLSHMIHTEHKQYLTDELVASTT 251

QY      249  GATTAAGDTTITYIAFAVNLNLRSPALTELVKAPGLMRNALDVLFDNLIRIGTVRF 308
Db      252  IVTLNAGHEATVHQIGNSVRIILSEGLDPKTLFHDETERTERTVEETLLIICAPVHI-FGRW 310

QY      309  ARQDLEYCGASIKKGEWVFLLI PSALRDGTVFSREPDVDVRDGTASLAYGKRGPHVCGV 368
Db      311  VLEPVEIDGVQFKQDKVSLTLAANLNDPAKFSDDPLAQOPRNEGANVSFGAGIHFCIGA 370

QY      369  SLATLEAEIANGVITFRPRPEMKLKETPVFG--YHAFNRISLNV 411
Db      371  PLARENLNALPLLFKRLPGKIAEPKPVKDYH--FGLGLEIDL 413

```

A:70707  
 cytochrome P450 RV0766c - Mycobacterium tuberculosis (strain H37Rv)  
 N:Contains: oxidoreductase (EC 1.-.-.-)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: A70707  
 R:Cole, S.T.; Broesch, R.; Parthill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500, MUID:98295987; PMID:9634230  
 A:Accession: A70707  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-402 <COL>  
 A:Cross-references: GB:280226; GB:AL123456; NID:g3261638; PIDN:CAB02396.1; PID:g1550644  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 A:Gene: RV0766c  
 C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
 C:Keywords: oxidoreductase  
 F:236-372/Domain: cytochrome P450 homology <CYP>  
 Query Match 21.2%; Score 450; DB 1; Length 402;  
 Best Local Similarity 37.4%; Pred. No. 3,5e-24;  
 Matches 128; Conservative 64; Mismatches 179; Indels 24; Gaps 9.

Oy 13 PAFDFKFAAGYADDPPEALERTLEATPIFYWDEGRSVLTRYHDSVAFRD-----ERF 67  
 8 PELVLDYVDYDFHDDPYFYRRLRDAENPLVNEENRNFVAVRHHVDVLCGFDDSTALSNAY 67  
 Oy 68 AVSSEEMESSLEVSASIAPELSDMKKYGFCGLPEEDHARVRLVNPSPFSRAIDLLRAEIQ 127  
 Db 68 GVS---LDPSRSTSEARVMS-----MLANDDPAHLRMRLTVSKGFPRRIRELEPVL 118  
 Oy 128 RTVQQLDASGCGEEFVVRDYAEGIPMRASIALTKLPACDECFRRRGSATYA-RALGVG 186  
 119 ELAIIHSDSALQTESPFDVAEFACKLPMDVISLIGPDTDRRAIRALADAVLRDEGVA 178  
 Oy 187 LVPOVDEETKTLVASVTEGALTLHDVLDERRRNPLENDVLTMLIOAEADGSRSLSTKELVA 246  
 Db 179 DVP-----PPMAASI-ELMRVYADLLIAEFRFRP-ANNTLSALIAAELDGCRLSDOEIYMA 231  
 Oy 247 LVGAIIAAGDTTAYTLAPAVLNLRSPEALELKAERGCLRNALDEVLREDNILRIGTV 306  
 232 FLPLMVAGNETTKLANNVYAAHHPGOLARFADBSRPLPMVEETLRKIDTSSQI-LA 290  
 Oy 307 RPARODLEYCGASIKKGEVFWLLIPSLRDGTVPSPDVPDDRDTGASL-AYGRGPHVC 365  
 Db 291 RTVAHDITLVYDTTIPGEVLLTLPGSANRRDVRDPPDQYRIGREIGCKLVSGSAHFC 350  
 Oy 366 PGVSLARLEAEIAVGTIFRRPEMKLKETPVGYN 400  
 Db 351 LGAHLARMEARVALGALLRRIRNRYEVDDNNVYRHN 385  
  
 RESULT 11  
 C87170  
 Probable cytochrome P450 (imported) - *Mycobacterium leprae*  
 C|Species: *Mycobacterium leprae*  
 C|Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 C|Accession: C87170  
 R|Cole, S.T.; Elgimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hamlin, N.; Davies, R.M.; Devlin, K.; Dutnoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.  
 Nature 409, 1007-1011, 2001  
 A|Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; A|Title: Massive gene decay in the leprosy bacillus.  
 A|Reference number: A86909; MUID:21128732; PMID:11234002  
 A|Accession: C87170  
 A|Status: preliminary



Db 225 ECGBPLTDEIATLVLTFTVAGNETTAHLIGNGVALLTHSDQLALLRSRRLLSQAVDE 284  
 Qy 294 VLRPNILRIGTVRFARODLEYCGASIKKGEWFLILPALRDGTVFSRDPVDFVARD-- 351  
 Db 285 LMRWCIPVQVOTQPRVATIEDLVGTVVRKGEQVAVITGAAGHPDRFPPEPFRDITRNHR 344  
 Qy 352 --TGASLAVGRPHVCPGVSLARLEAIAVGTIFRRFPPEMKLKEFP 395  
 Db 345 APHEAVHGFQGFHVCYCLGALAHQETVALMDTLFDRFSLALAVPP 390

## RESULT 14

S49051  
 Cytochrome P450 Cyt11 (EC 1.1.-.-) [validated] - Streptomyces fradiae (strain T59235)  
 C:Species: Streptomyces fradiae  
 A:Variety: strain T59235  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: S49051  
 R:Merion-Davies, L.A.; Cundliffe, E.  
 Mol. Microbiol. 13, 349-355, 1994  
 A:Title: Analysis of five tylosin biosynthetic genes from the Cyt11A region of the Streptomyces fradiae  
 A:Reference number: S49051; PMID:95075319; PMID:7984112  
 A:Accession: S49051  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-417 <MER>  
 A:Cross-References: EMBL:U08223; NID:G6849140; PIDN:AAA21341.1; PID:G473597  
 A:Experimental source: strain T59235  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994  
 C:Genetics:  
 A:Gene: Cyt11  
 C:Function:  
 A:Superfamily: involved in C20ring oxidation of O-mycaminoyl tylosone [validated, M011  
 C:Keywords: heme; oxidoreductase  
 F:253-388/Domain: cytochrome P450 homology <CYP>

Query Match 20.2%; Score 430; DB 1; Length 417;  
 Best Local Similarity 32.0%; Pred. No. 9.4e-23;  
 Matches 120; Conservative 70; Mismatches 161; Indels 24; Gaps 8;

Qy 34 RUREATPIYWDGR-SVLTTRYHDVSAVFRDERFAVSREWESSAAYSALP---ELSD 89  
 Db 41 RARAEAPVF-WDESRQAMQVFRY-----DDYLVSTNPOLFSSDFSPVPPELAI 91  
 Qy 90 MKKYGFL-GLPPEDHARVKLVNPSFTSRAIDLRAEIQRTVDQLLDASGQEPVVPVD 148  
 Db 92 LMGPGFGSIDPPRHGPKLVLSQAFTPRIATLEPRIAETRGGLDGLREKQIDVVD 151  
 Qy 149 VAEGIPMRATISALLKVAECDEKFRFGSATARALGVGLVPQVDEETKTLVASVTEGLAL 208  
 Db 152 LAVPLPVYIAELGIPADRDRLFREWVILNNEGMEFNP.LPDDSETMGARIKEMGY 211  
 Qy 209 LHDVLDERRRNP.LENDVLTMLLOAEGSRLSTKELVALVGAIIAAGTDTTLYIAFAVL 268  
 Db 212 LYRRIATLKRETPD-DLMSGLEAEVGRRLDDEIVNIVALLTFAGHISATLLGNLFL 270  
 Qy 269 NLRSPEALELYKAEGLMRNALDEVLR---FDNLRIGTVRFARODLEYCGASIKKGE 324  
 Db 271 VLDEHREAQELRADDLPGAIEETLYRSPFNIFRL-----LKEDTDIIIGHPMKAGQ 325  
 Qy 325 NVFLILPSALRDGTVFSRDPVDFVARDTASLAVGRPHVCPGVSLARLEAIAVGTIFR 384  
 Db 326 NVVAMIASANRDSAHSDPDTFVDRQPKHMSFGHGTHCISFLARLEAKVLELFED 385  
 Qy 385 RPEPMKLKETPVFGY 399  
 Db 386 EESDYRVEHDEVFEY 400

RESULT 15  
 T36526  
 probable cytochrome P450 hydroxylase - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Feb-2001  
 C:Accession: T36526  
 R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 Submitted to the EMBL Data Library, May 1999  
 A:Reference number: 221609  
 A:Accession: T36526  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-411 <M01>  
 A:Cross-References: EMBL:AL049754; PIDN:CAB42023.1; GSPDB:GN00070; SCODEB:SC10.14c  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCODEB:SC10.14c  
 C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology  
 C:Keywords: heme; iron; metalloprotein  
 F:250-385/Domain: cytochrome P450 homology <P45>  
 F:363/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 20.2%; Score 429; DB 2; Length 411;  
 Best Local Similarity 27.1%; Pred. No. 1.1e-22;  
 Matches 110; Conservative 93; Mismatches 191; Indels 12; Gaps 7;

Qy 13 PAFDFKPPAPGYAEDPPPAIERLEATPIFYWDEGRSVLTTRYHDVSAVFRDERFAVSRE 72  
 Db 13 PAFD--PMDPAFVADPYPAFAELRARGVLYYEPDQMLVPHHADVALLRDRRLGRTYQ 70  
 Qy 73 EHESSAAYSALP-----ELSDMKKYGFLGUPREHARVKLVNPSFTSRAIDLRAEQ 127  
 Db 71 HRFTHEDGRTPPPEQEPFHTLDHGM.LDEPPDHTIRRLVSKAFPTPTVERLKPYYH 130  
 Qy 128 RTVDQLLARSQGEFVDVVRDYAEGIPMRATISALLKVAECDEKFRFGSATARALGVGL 187  
 Db 131 GLADLVARLVAAGGDLITDYAEPLPVAIVAEMLGIP-ESDRALRPSAETICGM-YEL 188  
 Qy 188 VPQVDEETKTLVASVTEGLALHDVLDERRRNP.LENDVLTMLLOAEGSRLSTKELVAL 247  
 Db 189 NSEETAKAVAYASL-DESDYLRAIIARRKEP-GDDLISGLIAHDEDDRLTEQEMIST 246  
 Qy 248 VCAIIAAGTDTTLYIAFAVNLRLSPALAEVKAEPGLMRNALDEVLRFDNLRIGTVR 307  
 Db 247 CVLLNAGHEATVNTNGM.LALFRRHPOQLALRADHSLVPSAVELMKRYDTPQL-FER 305  
 Qy 308 FARODLEYCGASIKKGEWFLILPALRDGTVFSRDPVDFVARDTASLAVGRPHVCPG 367  
 Db 306 WLDIEIDITGLTPRGAVAM.LFGSANDPAVFTDPEELDLTRRNPHISFAGIHYCIG 365  
 Qy 368 VSLARLEAIAVGTIFRRFPPEMKLKETPVFGYHAFRNIESLNVIL 413  
 Db 366 APLARIELAASWTSLLKRAPGLRLAABERRRPNFVWRGLTELVEL 411

Search completed: October 2, 2003, 17:35:09  
 Job time: 15.5739 secs

**THIS PAGE BLANK (USPTO)**

52

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 2, 2003, 17:34:03 ; Search time 23.9103 seconds  
(without alignments)

2772.494 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124  
Sequence: 1 MTOEQANQSETPAFDFKPF.....HPAFNIESLVILKPSRAG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09C\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2124	100.0	419	US-10-014-717-8	Sequence 8, Appli
2	505.5	23.8	404	US-10-214-446-50	Sequence 50, Appli
3	498.5	23.5	392	US-10-214-446-32	Sequence 32, Appli
4	497.5	23.4	388	US-10-156-761-13776	Sequence 13776, A
5	495.5	23.3	408	US-10-214-446-4	Sequence 4, Appli
6	491	23.1	475	US-10-145-415-16	Sequence 16, Appli
7	488.5	23.0	475	US-10-145-415-22	Sequence 22, Appli
8	488	23.0	416	US-09-861-289-39	Sequence 39, Appli
9	488	23.0	416	US-09-860-846-39	Sequence 39, Appli
10	488	23.0	416	US-09-988-384B-39	Sequence 39, Appli
11	488	23.0	416	US-09-836-821-39	Sequence 39, Appli
12	486.5	22.9	470	US-10-145-415-6	Sequence 6, Appli
13	485.5	22.9	430	US-10-145-415-30	Sequence 30, Appli
14	484	22.8	416	US-09-793-708-18	Sequence 18, Appli
15	484	22.8	416	US-10-201-365-13	Sequence 13, Appli

16	484	22.8	416	12	US-10-160-539-18	Sequence 18, Appli
17	483.5	22.8	425	12	US-10-214-446-20	Sequence 20, Appli
18	483	22.7	399	15	US-10-156-761-9914	Sequence 9914, Ap
19	480.5	22.6	430	15	US-10-145-415-95	Sequence 95, Appli
20	475.5	22.4	430	15	US-10-145-415-8	Sequence 8, Appli
21	475.5	22.4	430	15	US-10-145-415-12	Sequence 12, Appli
22	475.5	22.4	430	15	US-10-145-415-20	Sequence 20, Appli
23	466	21.9	432	15	US-10-145-415-10	Sequence 10, Appli
24	463.5	21.8	430	15	US-10-145-415-32	Sequence 32, Appli
25	452.5	21.3	416	15	US-10-156-761-14997	Sequence 14997, A
26	452.5	21.3	430	15	US-10-145-415-18	Sequence 18, Appli
27	-451.5	21.3	429	15	US-10-145-415-14	Sequence 14, Appli
28	450	21.2	393	15	US-10-156-761-9525	Sequence 9525, Ap
29	450	21.2	405	15	US-10-156-761-12073	Sequence 12073, A
30	448	21.1	430	15	US-10-145-415-4	Sequence 4, Appli
31	448	21.1	430	15	US-10-145-415-28	Sequence 28, Appli
32	447	21.0	430	15	US-10-145-415-24	Sequence 24, Appli
33	447	21.0	457	15	US-10-156-761-11073	Sequence 11073, A
34	443.5	20.9	430	15	US-10-145-415-36	Sequence 36, Appli
35	443	20.9	399	12	US-10-314-657-37	Sequence 37, Appli
36	435.5	20.5	426	15	US-10-145-415-34	Sequence 34, Appli
37	435.5	20.5	430	15	US-10-145-415-2	Sequence 2, Appli
38	435	20.5	415	12	US-10-214-446-56	Sequence 56, Appli
39	433.5	20.4	399	15	US-10-156-761-7959	Sequence 7959, Ap
40	431	20.3	397	15	US-10-156-761-14715	Sequence 14715, A
41	428	20.2	415	12	US-10-214-446-46	Sequence 46, Appli
42	427.5	20.1	416	15	US-10-156-761-7958	Sequence 7958, Ap
43	425.5	20.0	402	15	US-10-205-032-8	Sequence 8, Appli
44	425	20.0	410	15	US-10-205-032-6	Sequence 6, Appli
45	418	19.7	404	12	US-10-214-446-40	Sequence 40, Appli

## ALIGNMENTS

RESULT 1  
US-10-014-717-8  
; Sequence 8, Application US/10014717  
; Publication No. US20020192778A1  
GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Liqon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/10/014,717  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US/09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 419  
; TYPE: prt  
; ORGANISM: Sorangium cellulosum  
US-10-014-717-8  
Query Match 100.0%, Score 2124, DB 14, Length 419;  
Best Local Similarity 100.0%, Pred. No. 2.2e-198;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTOEQANQSETPAFDFKPFAGVADDPFAIERREATPIFYWDEGRSWLTRHDVSA 60  
DB 1 MTOEQANQSETPAFDFKPFAGVADDPFAIERREATPIFYWDEGRSWLTRHDVSA 60  
QY 61 VRDRFPVSRWESSAAYSATPELSPMKKYGFGLPEDHARVRKLVNFSFTRAD 120  
DB 61 VRDRFPVSRWESSAAYSATPELSPMKKYGFGLPEDHARVRKLVNFSFTRAD 120  
QY 121 LTRAEIORTVQDLDRSGOEEFDVVRDYAEGIPKRAISALUKVPAECDEKFRFRGSATA 180

```
Db 121 LRAEIQRTVDQLLDARSQGEEDVVRDYAEIGIPMAISALLKVPACDEKFRFSGATA 180
Qy 181 RAIGVGLVPOVDEETKTLVAVTEGIALLDHVDERRRNPLENDVLTMLQAEADGRLS 240
Db 181 RAIGVGLVPOVDEETKTLVAVTEGIALLDHVDERRRNPLENDVLTMLQAEADGRLS 240
Qy 241 TKELVALVGAIIAAGDTTTLIAFAVNLNLSPEALEIKVABEGMLRNALDEVLPDNI 300
Db 241 TKELVALVGAIIAAGDTTTLIAFAVNLNLSPEALEIKVABEGMLRNALDEVLPDNI 300
Qy 301 LRIGVFRPARODLEYGASIKKGEVFLIPSLARDGTVESRPDVEDVRDGTASLAYGR 360
Db 301 LRIGVFRPARODLEYGASIKKGEVFLIPSLARDGTVESRPDVEDVRDGTASLAYGR 360
Qy 361 GPHVCGVSLARLEAEIAGVTIFRRPEMKLKETPVGYPHAPRNIESLNVILKPSKAG 419
Db 361 GPHVCGVSLARLEAEIAGVTIFRRPEMKLKETPVGYPHAPRNIESLNVILKPSKAG 419
```

## RESULT 2

```
US-10-214-446-50
; Sequence 50; Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-50
```

```
Query Match 23.8%; Score 505.5; DB 12; Length 404;
Best Local Similarity 32.5%; Pred. No. 1,4e-40;
Matches 133; Conservative 88; Mismatches 163; Indels 25; Gaps 12;
```

```
Qy 22 PGVAEDPEPAIERLEA--TFIPYWDGSGSWLTRYHDVAVPRDERFAVSREMESSAE 79
Db 5 PELRLDYPYQMRRETGRMHLSSANTGRWFLPHHADITLALRDEFSASRKAQGFVNQ 64
Qy 80 YSSAI-PELS---DMKKYGLFGLPEDHARVKLVNPSFTSRAIDILRAIORTVDQL 134
Db 65 FPAEVAPEFARFNEAISRWLVLDQPE-HRQLQMQGFTRLITMEKIKRQVCDLI 123
Qy 135 DA--RSQGEEDVVRDYAEIGIPMAISALLKVPACDEKFRFSGATARALGVLVPOVD 192
Db 124 DAEVKKGSTEF--MTEYAHPPPAKVIAMGVNPEDPAPVWSEDLNPAQ--SLRP-- 177
Qy 193 EETKTLVAVTEGIALLDH---VLDERRRNPLENDVLTMLQAEADGSLSTKELVALV 248
Db 178 --TLEMFRRAQDGLAMMDYFARLLPERREN--GDDVLSLLLSAESEGEWMTAEQVLANC 234
Qy 249 GATIAAGDTTTLIAFAVNLNLSPEALEIKVABEGMLRNALDEVLPDNIIRIGTVRF 308
Db 235 TQIIVAGHETTRNLVANGVELLRYPQORALLSREPLMSAVREIMRFESPIQF-IRRV 293
Qy 309 ARQDLEYGASIKKGEVFLIPSLARDGTVESRPDVEDVRDGTASLAYGRGPHVCPGV 368
Db 294 ARBDFEFGAIVEGDLVIMLGSANRDPAPFDDPTDILTRNPTGHLAFGQGWPHVCVGA 353
```

```
Qy 369 SLARLEAEIAGVTIFRRPEMKLK-ETPVGYHAPRNIESLNVILKPS 416
Db 354 ALAELEGVVSFRITLDRLPGLEIRTHEPERIRNPMILRGASJLDIGRES 402
```

## RESULT 3

```
US-10-214-446-32
; Sequence 32; Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-32
```

```
Query Match 23.5%; Score 498.5; DB 12; Length 392;
```

```
Best Local Similarity 33.8%; Pred. No. 6,6e-40;
Matches 127; Conservative 64; Mismatches 164; Indels 21; Gaps 5;
```

```
Qy 20 FAVGAEDPEPAIERLEATPIF--YWDGSGSWLTRYHDVAVPRDERFAVSREMESS 76
Db 9 YGADFTANPYPAKUREAGPVHEVMPPDGFQFWLVGHEGRALADRLA-----K 61
Qy 77 SAEYSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAIDILRAIORTVDQLLA 136
Db 62 SPEVIGVRPEEDIIICVHLIADADAPDHTLRRLVYGEFTGRVREGIRPRIQULTTELADA 121
Qy 137 RSQGEEDVVRDYAEIGIPMAISALLKVPACDEKFRFSGATARALGVLVPOVDETR 196
Db 122 MEPAGRADLVDAFAYPLIIVICELGVPAEDROTFRKNSQL-----VPTGEQF-- 172
Qy 197 TLVAVTEGIALLDHVDERRRNPLENDVLTMLQAEA-DGSRSLSTKELVALVGAIIAAG 255
Db 173 -FGQAVWDPAAYDALIEDKRAAGPTDILSLITARAEDGRLSGPELRAMAYLLILL 231
Qy 256 TDTTTLIAFAVNLNLSPEALEIKVABEGMLRNALDEVLPDNIIRIGTVFRPARODLEY 315
Db 232 HETTVMLIANTVNLITHEOIALARADPDLDGTIEELRYDGPVETGTFRTREAVTI 291
Qy 316 CGASIKKGEVFLIPSLARDGTVESRPDVEDVRDGTASLAYGRGPHVCPGVSLARLEA 375
Db 292 GGEIAGQIVLVGIALRDRPARFPDPRDIRRDTRGHLAFGHIHICLAPLARLBS 351
Qy 376 EIAVGTIFRRPEMKL 391
Db 352 RIARLTLDLRFPDLEL 367
```

## RESULT 4

```
US-10-156-761-13776
; Sequence 13776; Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
```

; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 13776  
 ; LENGTH: 388  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 ; US-10-156-761-13776

Query Match 23.4%; Score 497.5; DB 15; Length 388;  
 Best Local Similarity 34.0%; Pred. No. 8.1e-40;  
 Matches 136; Conservative 59; Mismatches 144; Indels 61; Gaps 10;

QY 21 APGVADDPFAIERREATPIFY---WBGKRWLTRYHDVSAVFRDERFAVSREEMESS 77  
 DB 4 AEGLLADYAYVDRLDTPAHPHRIAGTGKPMALVTRYDDV-----RE----- 46  
 QY 78 AEYSAPPELSDMKXYGFLG-----LPP-----EDHARVRKLVNPSFTSRAID 120  
 DB 47 ---GLANPLSLDKKHALPGNYRGALPPALDANLNDAPDHTRIRKVGAFPLRVE 103  
 QY 121 LLRAEIQRTVDQLDARSGOEFDVVDYAEIGPMRAISALIKVPAECDEKFRFGSATA 180  
 DB 104 QLRREVETARILDLGTHGSTDLIASYAAPLPITVICDLGVE---DEHRRD----- 155  
 QY 181 PALGVGV-----POVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQABA 234  
 DB 156 RAMTDPVLTVPDPARPDVARE-----SVVSLGFFFTGLADKRKNPAD-DLISDLIAVOE 208  
 QY 235 DGSRLSTKELVALVGAIIAAGTDTIYLAFVNLULRSPEALEVKAEPGLMRNALDEV 294  
 DB 209 EGDRIITBELMSLAPLIPAGENTVHLIGNVALLHHPQLALREDPARLPVASEF 268  
 QY 295 LRFNULIRIGTVRFARODLEYGASIKKGEWFFLLIPALRDGTFSRDPVFDVARDTGA 354  
 DB 269 ARYESPALAIRFPFVRVTIGGVVPAGEYTLISLSANRDPSPFPDPDLGLDQAAG 328  
 QY 355 SLAYGRGPHVCPGVSALAEIAVGTFRPPEKKLKT 394  
 DB 329 HLAIGHGVHVCIGAPLARLETEVALALAEPPDLALAE 368

RESULT 5  
 ; US-10-214-446-4  
 ; Sequence 4, Application US/10214446  
 ; Publication No. US20030180742A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weiner, David  
 ; APPLICANT: Burk, Mark J.  
 ; APPLICANT: Hitchman, Tim  
 ; APPLICANT: Pujol, Catherine  
 ; APPLICANT: Richardson, Toby  
 ; APPLICANT: Short, Jay M.  
 ; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING  
 ; FILE REFERENCE: 09010-500001  
 ; CURRENT APPLICATION NUMBER: US/10/214,446  
 ; CURRENT FILING DATE: 2002-08-05  
 ; PRIOR APPLICATION NUMBER: US 60/309,497  
 ; PRIOR FILING DATE: 2001-08-03  
 ; NUMBER OF SEQ ID NOS: 59  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4

; LENGTH: 408  
 ; TYPE: PRT  
 ; ORGANISM: Bacterial  
 ; US-10-214-446-4

Query Match 23.3%; Score 495.5; DB 12; Length 408;  
 Best Local Similarity 32.3%; Pred. No. 1.4e-39;  
 Matches 122; Conservative 72; Mismatches 157; Indels 27; Gaps 8;

QY 27 DDPFAIERRE-----ATPIFYWBGKRWLTRYHDVSAVFRDERFAVSREEMESSA 78  
 DB 21 NPYLFEIRHGVVQVRNLPVL-----EVMWGTGVDEAVAAALTDPLLS-SSPVGVNGL 74  
 QY 79 EYSAPPELSDMKXYGFLGPEEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDQLDARS 138  
 DB 75 EEBMAHQERTIVLWASMLVANGEDHTRIRNLVSAFTARVEQLAPRQAHTDALDVA 134  
 QY 139 QOEFDVVDYAEIGPMRAISALIKVPAECDEKFRFGSATAHALGVGV---POVDEETK 196  
 DB 135 ARGSDLVSEFALPLPMVLSDLIGIPAEQPDPAF-----LAVGLIMPNTPERLA 186  
 QY 197 TLVASVTEGLALHDVLDERRRNPLENDVLTMLQABDGSRLSTKELVALVGAIIAAGT 256  
 DB 187 KGAPRAELTFEEPLINQRKEP-KDDLALCAQRE-ERISDRLEITAMTILTLTLAGH 244  
 QY 257 DTTIYLAFVNLULRSPEALEVKAEPGLMRNALDEVLPDNLIRIGTVRFARODLEYC 316  
 DB 245 ETTASLIANGVHALRHREQFATLRDPSLLPGAILBELRTEGVPVSRGVAFPTDPYBIG 304  
 QY 317 GASIKKGEWFFLLIPALRDGTFSRDPVFDV-RRDGTASLAYGRGPHVCPGVSALARLEA 375  
 DB 305 GVTVPAGEMIIIGLAAANRDPAHYDRPDILDAARREVPOQLAFGHVFCIGAPLARABA 364  
 QY 376 EIAVGTFRPPEKKLKT 393  
 DB 365 RIAIGTLIRPPDLRLAD 382

RESULT 6  
 ; US-10-145-415-16  
 ; Sequence 16, Application US/10145415  
 ; Publication No. US20030068788A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Buckel, Thomas Gunter  
 ; APPLICANT: Hammer, Philip Eugene  
 ; APPLICANT: Hill, Dwight Steven  
 ; APPLICANT: Ligon, James Madison  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Pachtako, Johannes Paul  
 ; APPLICANT: Zitzke, Ross Eric  
 ; TITLE OF INVENTION: Methods and Compositions for Making Emamectin  
 ; FILE REFERENCE: SYN-117 109846,312  
 ; CURRENT APPLICATION NUMBER: US/10/145,415  
 ; CURRENT FILING DATE: 2002-05-14  
 ; PRIOR APPLICATION NUMBER: US 60/291,149  
 ; PRIOR FILING DATE: 2001-05-16  
 ; NUMBER OF SEQ ID NOS: 105  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16  
 ; LENGTH: 475  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces albofaciens  
 ; US-10-145-415-16

Query Match 23.1%; Score 491; DB 15; Length 475;  
 Best Local Similarity 29.6%; Pred. No. 4.8e-39;  
 Matches 129; Conservative 83; Mismatches 156; Indels 68; Gaps 11;

QY 6 ANOSETKPADEFKFAFGYA-----EDDPFAIERREATPIFY---YMEG 47  
 DB 34 ASPTTSPATTDRTTLTSYGLHGPENLMEPELLDNYTGYGTLREDAPLVRARFIDS 93  
 QY 48 RSWLTRYHDVSAVFRDERFAVSREEMESSAEYSAPPELS-----DMKYGGLGFLP--- 99





```

RESULT 9
US-09-860-846-39
; Sequence 39, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRF
; ORGANISM: Streptomyces venezuelae
US-09-860-846-39

```

```

Query Match      23.0%; Score 488; DB 10; Length 416;
Best Local Similarity 30.5%; Pred. No. 7.7e-39;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

```

```

QY      1 MTOEQANQSEKPAEDFKFPAAGYADDPFAIERLEATPI--FYWDEG-RSMVLTTRYD 57
      1 MKRTQGGTTASPPVLDLGLGODFADPYPTAARLRAEGPARVRTPEDDEVWLVGXYDR 60
DB      1
QY      58 VSAVFRDERFAVSREWESSAESSAIPELSDMKKYGFLGPPEDHARVRKLVNPSFTSR 117
      61 ARAVLADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPRHTRLRKLVAREFTWR 114
DB      118 AIDLRAEIQRTVDQLLDARSQGE--FDVVDYAEIGIPMRAISALLKVPACDEKFRFP 175
      115 RVELLRPRVOEIVDGLVDMMLAAPGRADLMESLAMPDITVISELGVPEPDRAAFRW 174
DB      176 GSATRALGVGLVPQVDEETKTLVASVTGALLHDVLDERRRNPENDVLTMLQAEND 235
      175 TDAF-----VFPPDPAQAOTAMAMESGYLSRL--IDSKRGQGEEDLLSLVRTSD 224
DB      236 GSRLTKEIVALVGAIIAAGTDTTYLIAFAVLNLRSPDEALVKAEGMLRNALDEV 295
      225 GSRLTSEELGMAHILVAGHETTVNLANGMYALLSHPDQALALRADWTLTDGAVEEW 284
QY      296 REDNLRIGTVAFARODLEYCGASIKKGMVFLIPSLARDGVFSRPVDPVRRDGTAS 355
      285 RYEGPVESATYRFPVPEVDLDTVLPAGDTVLVLADAHRTBERPDPFRFDIRRDTAGH 344
DB      356 LAYGGRPHVCPGVSLARLEAEIAGVTIFRRPEMKIKETP---VFGYHAPFNIESLV 411
      345 LAFGHGIFHCIGAPLARLEARIAVRALBERCDLADVSPGELVWYPNMIGLKALPLI 403
QY      345

```

```

RESULT 10
US-09-988-384B-39
; Sequence 39, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537

```

```

; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRF
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-39

```

```

Query Match      23.0%; Score 488; DB 11; Length 416;
Best Local Similarity 30.5%; Pred. No. 7.7e-39;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

```

```

QY      1 MTOEQANQSEKPAEDFKFPAAGYADDPFAIERLEATPI--FYWDEG-RSMVLTTRYD 57
      1 MKRTQGGTTASPPVLDLGLGODFADPYPTAARLRAEGPARVRTPEDDEVWLVGXYDR 60
DB      1
QY      58 VSAVFRDERFAVSREWESSAESSAIPELSDMKKYGFLGPPEDHARVRKLVNPSFTSR 117
      61 ARAVLADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPRHTRLRKLVAREFTWR 114
DB      118 AIDLRAEIQRTVDQLLDARSQGE--FDVVDYAEIGIPMRAISALLKVPACDEKFRFP 175
      115 RVELLRPRVOEIVDGLVDMMLAAPGRADLMESLAMPDITVISELGVPEPDRAAFRW 174
QY      176 GSATRALGVGLVPQVDEETKTLVASVTGALLHDVLDERRRNPENDVLTMLQAEND 235
      175 TDAF-----VFPPDPAQAOTAMAMESGYLSRL--IDSKRGQGEEDLLSLVRTSD 224
DB      236 GSRLTKEIVALVGAIIAAGTDTTYLIAFAVLNLRSPDEALVKAEGMLRNALDEV 295
      225 GSRLTSEELGMAHILVAGHETTVNLANGMYALLSHPDQALALRADWTLTDGAVEEW 284
QY      296 REDNLRIGTVAFARODLEYCGASIKKGMVFLIPSLARDGVFSRPVDPVRRDGTAS 355
      285 RYEGPVESATYRFPVPEVDLDTVLPAGDTVLVLADAHRTBERPDPFRFDIRRDTAGH 344
DB      356 LAYGGRPHVCPGVSLARLEAEIAGVTIFRRPEMKIKETP---VFGYHAPFNIESLV 411
      345 LAFGHGIFHCIGAPLARLEARIAVRALBERCDLADVSPGELVWYPNMIGLKALPLI 403
QY      345

```

```

RESULT 11
US-09-836-821-39
; Sequence 39, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRF
; ORGANISM: Streptomyces venezuelae
US-09-836-821-39

```

```

Query Match      23.0%; Score 488; DB 11; Length 416;
Best Local Similarity 30.5%; Pred. No. 7.7e-39;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

```

```

QY      1 MTOEQANQSEKPAEDFKFPAAGYADDPFAIERLEATPI--FYWDEG-RSMVLTTRYD 57
      1 MKRTQGGTTASPPVLDLGLGODFADPYPTAARLRAEGPARVRTPEDDEVWLVGXYDR 60
DB      1
QY      58 VSAVFRDERFAVSREWESSAESSAIPELSDMKKYGFLGPPEDHARVRKLVNPSFTSR 117

```

```

Db 61 ARAVALDPRFS---KOWRNS---TTPLEBAALNNMLESDDPRHRLAKLVAREFTWR 114
QY 118 AIDLRAEIQRTYDOLLDRASGOEE--FDVVRDYAGCIIPKRAISALKVPAECDEKFRFR 175
Db 115 RVELLPRVQEIYDGLVDMLAAPDRADIMESLAMPITVISELIGVEEDPRAAFRW 174
QY 176 GSATARALGVLPQVDEEFTKLVASTEGTALLHVDLDERRRNPLENDVLTMLQAEAD 235
Db 175 TDHF-----VFPDDPAQOTMAEMSGYLSRL---IDSKRODGEDLLSALVTSDED 224
QY 236 GSRLLSTKELVALGAIIAAGTDTTIVLIAFAVNLILRSPEALEVVAEPGLKRNALDEV 295
Db 225 GSRLLSEELLGMHILLVAGHETTVMILANGMYALLSHPDQLAALRADMTLLDGAVEML 284
QY 296 RPNILIRIGVRARADOLEYCGASIKKGEVFLILISALRDGVFSRPDVPDRDPTGAS 355
Db 285 RYEGPESATYRRPVPVDDGTIVIPAGDVLVVLADAHETPERFPDPHFRDRIRODAGH 344
QY 356 LAYGRGPHVCPGVSLARLEAEIAGVTIFRRPEPMKLETP---VEGYHAPAFRNIESINV 411
Db 345 LAFGHGHIHCTGAPLARLEARIIVRALLEKCPDLADVSGLVWIPWPIRGKALPI 403

RESULT 12
US-10-145-415-6
; Sequence 6, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zirkle, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Enamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; PRIORITY FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Streptomyces timosus
US-10-145-415-6

Query Match 22.9%; Score 486.5; DB 15; Length 470;
Best Local Similarity 30.6%; Pred. No. 1.3e-38;
Matches 124; Conservative 81; Mismatches 147; Indels 53; Gaps 10;

QY 22 PGYAEPPFAIERLEATPIF---YWDGRSWLTRYHVSASFRRERFAVSEBMESSA 78
Db 60 PELLENPYTGTLRREGAPLVARAFIDSDPIWLTREVVRDORF-----V 109
QY 79 EYSSAIPELSDMKK---YGLFGLP-----PEDHARVRLVNSFSTRSA 118
Db 110 NNPTLVPGIADXPBARLLELFGIPEDLAPYLTDNLITSDPPDHTLRRLVSRAPFARR 169
QY 119 IDLLRAEIQRTYDOLLDRASGOEE--FDVVRDYAGCIIPKRAISALKVPAECDEKFRFR 175
Db 170 IQDLRPRVERITBELERLPDHAEDGVVDLVEHFAVPLPTVICELVIGIDEDRALMRFP 229
QY 176 GSATARALGVLPQVDEEFTKLVASTEGTALLHVDLDERRRNPLENDVLTMLQAEAD 235
Db 230 GADLA-----SLNP-----KRIGATPEMISHHEIDERRA-ALADDLSGLIRKODD 277
QY 235 DGSRLSTKELVALGAIIAAGTDTTIVLIAFAVNLILRSPEALEVVAEPGLKRNALDEV 294
Db 278 DGRLLSDVENMTVLTLVLGHETTAAHLISNGTIALTLTHPDORRLIDEDBALLPRAVHEL 337

```

```

QY 295 LRFNLIIRIGTVARADOLEYCGASIKKGEVFLILISALRDGVFSRPDVPDRDPTGA 354
Db 338 MWCCEGIQTOURLRYALDEDEVAGVQVROGEALMFLSVAAHNHPRHVTGERLDLTRQPG 397
QY 355 ---SLAYGRGPHVCPGVSLARLEAEIAGVTIFRRPEPMKLETP 395
Db 398 RAEDHVGFGHMYCIGASLARQEAFAVAYGKLLTRFYDLALALTP 442

RESULT 13
US-10-145-415-30
; Sequence 30, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zirkle, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Enamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; PRIORITY FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Streptomyces lydicus
US-10-145-415-30

Query Match 22.9%; Score 485.5; DB 15; Length 430;
Best Local Similarity 31.4%; Pred. No. 1.4e-38;
Matches 129; Conservative 81; Mismatches 150; Indels 51; Gaps 13;

QY 22 PGYAEPPFAIERLEATPIF---YWDGRSWLTRYHVSASFRRERFAVSEBMESSA 78
Db 25 PALIGDPFAGYGALRQGVVVRGRFVDDSPVWFVTFEEVREVLARDORF---RNPVSSA 81
QY 79 EYSSAIPELSDMKK---YGLFGLPPE-----DHARVRLVNSFSTRAIDL 122
Db 82 --PDADPEDTPLSRLLDMMGFPBHLRVYLIGSLTUNNADPDHTRRLVSRAPFARKITDL 139
QY 123 RAEIQRTYDOLLDRASGOEE--FDVVRDYAGCIIPKRAISALKVPAECDEKFRFGSAT 179
Db 140 RPRVAGIADLELRLARLEHAEEDGVVDLIGHFAVPLPTVICELVIGIDEDRALMRFP 199
QY 180 ARAVGGLVLPQVDEEFTKLVASTEGTALLHVDLDERRRNPLENDVLTMLQAEADGSR 238
Db 200 ----SLNP-----DRMRSRSPAMIDH---IHETIAARR-ALTDLLSELIRTHDDGSR 247
QY 239 LSTKELVALGAIIAAGTDTTIVLIAFAVNLILRSPEALEVVAEPGLKRNALDEVLRPD 298
Db 248 LSVENVMTVLTVVLGHETTAAHLISNGTIALTLTHPDORRLIDEDBALLPRAVHELMRC 307
QY 299 NILIRIGTVARADOLEYCGASIKKGEVFLILISALRDGVFSRPDVPDRDPTGA---- 354
Db 308 GPVHMTQLRYAARDVELAGVRIKGDVAVQILIVSANRDPRHNYTEPRLDLTRHPAGHAEN 367
QY 355 SLAYGRGPHVCPGVSLARLEAEIAGVTIFRRPEPMK-----LKEPVPFG 398
Db 368 HVGFGHAGHAYCIGATLAKOEGEVALGALLRHFPPELSLAAPDALERTPVPG 418

RESULT 14
US-09-793-708-18
; Sequence 18, Application US/09793708
; Publication No. US20030104597A1

```

```

; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, LI
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PR
; ORGANISM: Streptomyces venezuelae
US-09-793-708-18

```

```

Query Match      22.8%; Score 484; DB 11; Length 416;
Best Local Similarity 30.6%; Pred. No. 1.9e-38;
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQETKAPDFKFPAGYADPPAIRLRATPI--FYWDEG-RSMVLTIRYHDVAV 61
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 QQGTASPPVLDLGLGODFADPPYPTVARLRAGPAHRVTRTEGDEWLVVGYDRARAV 64
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 FDERFAVSREEMESSAEYSSAIPELSDMKYKGLFGLPPEDHARVRLVNPSTSAIDL 121
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 LADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPRTRKLVARERTMRVRL 118
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 LRAEIQTVDOLLARSQOE--FDVVRDYAEGIPMRASIALIKVPAECDEKFRFGSAT 179
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 LRPVQEIYDGLVDAMLAPDGRADLMESLWPLPTVISELGVPEPRAAFRWTDAP 178
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 AALGVGLVPOVDEETKTLVASVTBGLALHVDLERRRNPLENDVLTMLQAEADGSL 239
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 -----VFPDDPAQAQTAMAEWSGYLSRL--IDSKRGQDGEDLLSALVRTSDEGSL 228
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 STKELVALVGATIIAAGTDTTIVLIAFAVNLRSPEALVKAEPGLMKNALDEVLRPN 299
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 229 TSEELGMAHILLVAGHETTVLIANGMTALSHPDQALALADMTLLDGAVEEMLRVYG 288
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 ILRIGTVRPARODLEYCGASIKKGEWFLIPSAIRDGTVFSRPDVFDVRRDGTASLAYG 359
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 289 PVESATYRPPVPEVDLDGTIVIPAGDTVLVLAADARTPERFPDPRFDIRRDTAGHLAFG 348
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 RGPHYCPGVSLARLEAIVAGTIFRRFPBMKLETP---VFGYHAPFRNIESLNV 411
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 HGIIHFCIGAPLARLEARIARIALRLLERCPDLALDVSPGELVWYPNPIMIRGLKALPI 403
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

US-10-201-365-13
; Sequence 13, Application US/10201365
; Publication No. US20030148469A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, LI

```

```

; TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A MODULAR
; FILE REFERENCE: 300622002103
; CURRENT APPLICATION NUMBER: US/10/201,365
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 416
; TYPE: PR
; ORGANISM: Streptomyces venezuelae
US-10-201-365-13

```

```

Query Match      22.8%; Score 484; DB 12; Length 416;
Best Local Similarity 30.6%; Pred. No. 1.9e-38;
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQETKAPDFKFPAGYADPPAIRLRATPI--FYWDEG-RSMVLTIRYHDVAV 61
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 QQGTASPPVLDLGLGODFADPPYPTVARLRAGPAHRVTRTEGDEWLVVGYDRARAV 64
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 FDERFAVSREEMESSAEYSSAIPELSDMKYKGLFGLPPEDHARVRLVNPSTSAIDL 121
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 LADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPRTRKLVARERTMRVRL 118
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 LRAEIQTVDOLLARSQOE--FDVVRDYAEGIPMRASIALIKVPAECDEKFRFGSAT 179
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 LRPVQEIYDGLVDAMLAPDGRADLMESLWPLPTVISELGVPEPRAAFRWTDAP 178
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 AALGVGLVPOVDEETKTLVASVTBGLALHVDLERRRNPLENDVLTMLQAEADGSL 239
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 -----VFPDDPAQAQTAMAEWSGYLSRL--IDSKRGQDGEDLLSALVRTSDEGSL 228
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 STKELVALVGATIIAAGTDTTIVLIAFAVNLRSPEALVKAEPGLMKNALDEVLRPN 299
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 229 TSEELGMAHILLVAGHETTVLIANGMTALSHPDQALALADMTLLDGAVEEMLRVYG 288
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 ILRIGTVRPARODLEYCGASIKKGEWFLIPSAIRDGTVFSRPDVFDVRRDGTASLAYG 359
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 289 PVESATYRPPVPEVDLDGTIVIPAGDTVLVLAADARTPERFPDPRFDIRRDTAGHLAFG 348
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 RGPHYCPGVSLARLEAIVAGTIFRRFPBMKLETP---VFGYHAPFRNIESLNV 411
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 HGIIHFCIGAPLARLEARIARIALRLLERCPDLALDVSPGELVWYPNPIMIRGLKALPI 403
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: October 2, 2003, 17:47:43  
Job time : 25.9103 secs

**THIS PAGE BLANK (USPTO)**



QY 361 GPHVCGVSLARLEAIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419  
 Db 361 GPHVCGVSLARLEAIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419

# RESULT 2

US-09-413-814-71  
 ; Sequence 71, Application US/09413814  
 ; Patent No. 6223064  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
 ; APPLICANT: Bristol-Myers Squibb, Co.  
 ; APPLICANT: Bayer, Stefan  
 ; APPLICANT: Bioecker, Helmut  
 ; APPLICANT: Brandt, Petra  
 ; APPLICANT: Cino, Paul M  
 ; APPLICANT: Dougherty, Brian A  
 ; APPLICANT: Goldberg, Steven L  
 ; APPLICANT: Hofle, Gerhard  
 ; APPLICANT: Mueller, Joachim  
 ; APPLICANT: Reichenbach, Hans  
 ; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
 ; FILE REFERENCE: PCT/US 99/23535  
 ; CURRENT APPLICATION NUMBER: US/09/413,814  
 ; EARLIER FILING DATE: 1999-10-07  
 ; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 71  
 ; LENGTH: 419  
 ; TYPE: PR1  
 ; ORGANISM: Sorangium cellulosum  
 ; US-09-413-814-71

Query Match 100.0%; Score 2124; DB 3; Length 419;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-210;  
 Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAPDFKFPAGYADPPPAIRLREATPIFYWDEGRSMLTRYHDVSA 60  
 Db 1 MTOEQANQSETKPAPDFKFPAGYADPPPAIRLREATPIFYWDEGRSMLTRYHDVSA 60  
 QY 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFLGPPEDHARVRLVNSFTSRAID 120  
 Db 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFLGPPEDHARVRLVNSFTSRAID 120  
 QY 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIGIPMAISALLKVPACDEKRRFGSATA 180  
 Db 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIGIPMAISALLKVPACDEKRRFGSATA 180  
 QY 181 RALGVLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGRS 240  
 Db 181 RALGVLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGRS 240  
 QY 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLLRSPEALELVKABPGIMRNALDEVLRDNI 300  
 Db 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLLRSPEALELVKABPGIMRNALDEVLRDNI 300  
 QY 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTFSRPDVVDVDRDGTASLAYGR 360  
 Db 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTFSRPDVVDVDRDGTASLAYGR 360  
 QY 361 GPHVCGVSLARLEAIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419  
 Db 361 GPHVCGVSLARLEAIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419

RESULT 3  
 US-09-568-102-8  
 ; Sequence 8, Application US/09568102

; Patent No. 6346404  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Zirkle, Rose  
 ; APPLICANT: Cyr, Devon  
 ; APPLICANT: Goerlach, Joern  
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 ; FILE REFERENCE: 4-30582A  
 ; CURRENT APPLICATION NUMBER: US/09/568,102  
 ; PRIOR FILING DATE: 2000-05-10  
 ; PRIOR APPLICATION NUMBER: 09/335,409  
 ; PRIOR FILING DATE: 1999-06-17  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 8  
 ; LENGTH: 419  
 ; TYPE: PR1  
 ; ORGANISM: Sorangium cellulosum  
 ; US-09-568-102-8

Query Match 100.0%; Score 2124; DB 4; Length 419;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-210;  
 Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAPDFKFPAGYADPPPAIRLREATPIFYWDEGRSMLTRYHDVSA 60  
 Db 1 MTOEQANQSETKPAPDFKFPAGYADPPPAIRLREATPIFYWDEGRSMLTRYHDVSA 60  
 QY 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFLGPPEDHARVRLVNSFTSRAID 120  
 Db 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFLGPPEDHARVRLVNSFTSRAID 120  
 QY 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIGIPMAISALLKVPACDEKRRFGSATA 180  
 Db 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIGIPMAISALLKVPACDEKRRFGSATA 180  
 QY 181 RALGVLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGRS 240  
 Db 181 RALGVLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGRS 240  
 QY 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLLRSPEALELVKABPGIMRNALDEVLRDNI 300  
 Db 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLLRSPEALELVKABPGIMRNALDEVLRDNI 300  
 QY 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTFSRPDVVDVDRDGTASLAYGR 360  
 Db 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTFSRPDVVDVDRDGTASLAYGR 360  
 QY 361 GPHVCGVSLARLEAIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419  
 Db 361 GPHVCGVSLARLEAIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419

RESULT 4  
 US-09-567-969-8  
 ; Sequence 8, Application US/09567969  
 ; Patent No. 635457  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Zirkle, Rose  
 ; APPLICANT: Cyr, Devon  
 ; APPLICANT: Goerlach, Joern  
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
 ; FILE REFERENCE: 4-30582A  
 ; CURRENT APPLICATION NUMBER: US/09/567,969  
 ; PRIOR FILING DATE: 2000-05-10  
 ; PRIOR APPLICATION NUMBER: 09/335,409  
 ; PRIOR FILING DATE: 1999-06-17  
 ; NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 419  
TYPE: PRF  
ORGANISM: Sorangium cellulosum  
US-09-567-969-8

Query Match 100.0%; Score 2124; DB 4; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.2e-210;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEANOSEKPAEDFKFPAQYADPPPAIERLEATPIFYWDEGSMVLTTRYHDSA 60  
DB 1 MTOEANOSEKPAEDFKFPAQYADPPPAIERLEATPIFYWDEGSMVLTTRYHDSA 60  
QY 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120  
DB 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120  
QY 121 LIRAEIORTVDQLDARSQOEFDVRYAEGIPMRASALKVPAECDEKRRFGSATA 180  
DB 121 LIRAEIORTVDQLDARSQOEFDVRYAEGIPMRASALKVPAECDEKRRFGSATA 180  
QY 181 RALGVLVPOVDEETKTIVASTEGIALHVDLDRRRNPLENDVLTMLQAEADGSRIS 240  
DB 181 RALGVLVPOVDEETKTIVASTEGIALHVDLDRRRNPLENDVLTMLQAEADGSRIS 240  
QY 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALELVKAPGLMRNALDEVLRFDNI 300  
DB 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALELVKAPGLMRNALDEVLRFDNI 300  
QY 301 LRIGTVRPARODLEFCGASIKKGEVFLIIPALRDGTFSRPDVFVARDTGASLAYGR 360  
DB 301 LRIGTVRPARODLEFCGASIKKGEVFLIIPALRDGTFSRPDVFVARDTGASLAYGR 360  
QY 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKEPVFGYHPAFRNIESLNVILKPSKAG 419  
DB 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKEPVFGYHPAFRNIESLNVILKPSKAG 419

RESULT 5  
US-09-568-480-8  
Sequence 8, Application US/09568480  
Patent No. 6355458  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOCHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,480  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 419  
TYPE: PRF  
ORGANISM: Sorangium cellulosum  
US-09-568-480-8

Query Match 100.0%; Score 2124; DB 4; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.2e-210;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEANOSEKPAEDFKFPAQYADPPPAIERLEATPIFYWDEGSMVLTTRYHDSA 60  
DB 1 MTOEANOSEKPAEDFKFPAQYADPPPAIERLEATPIFYWDEGSMVLTTRYHDSA 60

QY 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120  
DB 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120  
QY 121 LIRAEIORTVDQLDARSQOEFDVRYAEGIPMRASALKVPAECDEKRRFGSATA 180  
DB 121 LIRAEIORTVDQLDARSQOEFDVRYAEGIPMRASALKVPAECDEKRRFGSATA 180  
QY 181 RALGVLVPOVDEETKTIVASTEGIALHVDLDRRRNPLENDVLTMLQAEADGSRIS 240  
DB 181 RALGVLVPOVDEETKTIVASTEGIALHVDLDRRRNPLENDVLTMLQAEADGSRIS 240  
QY 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALELVKAPGLMRNALDEVLRFDNI 300  
DB 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALELVKAPGLMRNALDEVLRFDNI 300  
QY 301 LRIGTVRPARODLEFCGASIKKGEVFLIIPALRDGTFSRPDVFVARDTGASLAYGR 360  
DB 301 LRIGTVRPARODLEFCGASIKKGEVFLIIPALRDGTFSRPDVFVARDTGASLAYGR 360  
QY 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKEPVFGYHPAFRNIESLNVILKPSKAG 419  
DB 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKEPVFGYHPAFRNIESLNVILKPSKAG 419

RESULT 6  
US-09-568-486-8  
Sequence 8, Application US/09568486  
Patent No. 6355459  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOCHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,486  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1998-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 419  
TYPE: PRF  
ORGANISM: Sorangium cellulosum  
US-09-568-486-8

Query Match 100.0%; Score 2124; DB 4; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.2e-210;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEANOSEKPAEDFKFPAQYADPPPAIERLEATPIFYWDEGSMVLTTRYHDSA 60  
DB 1 MTOEANOSEKPAEDFKFPAQYADPPPAIERLEATPIFYWDEGSMVLTTRYHDSA 60  
QY 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120  
DB 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120  
QY 121 LIRAEIORTVDQLDARSQOEFDVRYAEGIPMRASALKVPAECDEKRRFGSATA 180  
DB 121 LIRAEIORTVDQLDARSQOEFDVRYAEGIPMRASALKVPAECDEKRRFGSATA 180  
QY 181 RALGVLVPOVDEETKTIVASTEGIALHVDLDRRRNPLENDVLTMLQAEADGSRIS 240  
DB 181 RALGVLVPOVDEETKTIVASTEGIALHVDLDRRRNPLENDVLTMLQAEADGSRIS 240  
QY 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALELVKAPGLMRNALDEVLRFDNI 300  
DB 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALELVKAPGLMRNALDEVLRFDNI 300



QY 301 LRIGTVFARODLEYCGASIKKGEWVFLIPSLARDGVSRPDVDRDTCASLAYS 360  
DB 301 LRIGTVFARODLEYCGASIKKGEWVFLIPSLARDGVSRPDVDRDTCASLAYS 360  
QY 361 GPHVCPGVSLARLEAEIAGVTIFRRFPPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419  
DB 361 GPHVCPGVSLARLEAEIAGVTIFRRFPPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419

## RESULT 7

US-09-568-472-8  
Sequence 8, Application US/09568472  
Patent No. 6358719  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,472  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 8  
LENGTH: 419  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-568-472-8

Query Match 100.0%; Score 2124; DB 4; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.2e-210;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAFDKFPAPGYADPPPAIERLRLEATPIFYWDEGRSWLTRHYDVA 60  
DB 1 MTOEQANQSETKPAFDKFPAPGYADPPPAIERLRLEATPIFYWDEGRSWLTRHYDVA 60  
QY 61 VFDERFAVSREWESSAEYSASIPELSDMKKYGGLPPEPHARVRLKVNPSFTSRAID 120  
DB 61 VFDERFAVSREWESSAEYSASIPELSDMKKYGGLPPEPHARVRLKVNPSFTSRAID 120  
QY 121 LTRAEIQRIVDQLLDARSQGEEDVVDYAEIGIPMRASIALLKVPACDEKFRFSGATA 180  
DB 121 LTRAEIQRIVDQLLDARSQGEEDVVDYAEIGIPMRASIALLKVPACDEKFRFSGATA 180  
QY 181 RALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLQAEADSGRLS 240  
DB 181 RALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLQAEADSGRLS 240  
QY 241 TKELVALVGAIIAAGDTTITYLAFVNLNLSPEALEVKAEGLMRNALDEVLRFDNI 300  
DB 241 TKELVALVGAIIAAGDTTITYLAFVNLNLSPEALEVKAEGLMRNALDEVLRFDNI 300  
QY 301 LRIGTVFARODLEYCGASIKKGEWVFLIPSLARDGVSRPDVDRDTCASLAYS 360  
DB 301 LRIGTVFARODLEYCGASIKKGEWVFLIPSLARDGVSRPDVDRDTCASLAYS 360  
QY 361 GPHVCPGVSLARLEAEIAGVTIFRRFPPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419  
DB 361 GPHVCPGVSLARLEAEIAGVTIFRRFPPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419

## RESULT 8

US-09-567-899-8  
Sequence 8, Application US/09567899  
Patent No. 6383787  
GENERAL INFORMATION:

APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/567,899  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 8  
LENGTH: 419  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-567-899-8

Query Match 100.0%; Score 2124; DB 4; Length 419;  
Best Local Similarity 100.0%; Pred. No. 1.2e-210;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAFDKFPAPGYADPPPAIERLRLEATPIFYWDEGRSWLTRHYDVA 60  
DB 1 MTOEQANQSETKPAFDKFPAPGYADPPPAIERLRLEATPIFYWDEGRSWLTRHYDVA 60  
QY 61 VFDERFAVSREWESSAEYSASIPELSDMKKYGGLPPEPHARVRLKVNPSFTSRAID 120  
DB 61 VFDERFAVSREWESSAEYSASIPELSDMKKYGGLPPEPHARVRLKVNPSFTSRAID 120  
QY 121 LTRAEIQRIVDQLLDARSQGEEDVVDYAEIGIPMRASIALLKVPACDEKFRFSGATA 180  
DB 121 LTRAEIQRIVDQLLDARSQGEEDVVDYAEIGIPMRASIALLKVPACDEKFRFSGATA 180  
QY 181 RALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLQAEADSGRLS 240  
DB 181 RALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLQAEADSGRLS 240  
QY 241 TKELVALVGAIIAAGDTTITYLAFVNLNLSPEALEVKAEGLMRNALDEVLRFDNI 300  
DB 241 TKELVALVGAIIAAGDTTITYLAFVNLNLSPEALEVKAEGLMRNALDEVLRFDNI 300  
QY 301 LRIGTVFARODLEYCGASIKKGEWVFLIPSLARDGVSRPDVDRDTCASLAYS 360  
DB 301 LRIGTVFARODLEYCGASIKKGEWVFLIPSLARDGVSRPDVDRDTCASLAYS 360  
QY 361 GPHVCPGVSLARLEAEIAGVTIFRRFPPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419  
DB 361 GPHVCPGVSLARLEAEIAGVTIFRRFPPEMKLKETPVFGYHPAFRNIESLNVILKPSKAG 419

## RESULT 9

US-09-105-537-39  
Sequence 39, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 39  
LENGTH: 416  
TYPE: PRT  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-39

Query Match 23.0%; Score 488; DB 3; Length 416;  
Best Local Similarity 30.5%; Pred. No. 9.4e-42;  
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

QY 1 MTOEQANSETPAPDFKFPAGYADDPFAIRLEATPI--FYWDEG-RSWLTRHYD 57  
DB 1 MRRTOGTTASPPVLDLGLGDFADPPYTAARLAEAGPAHRVTRPEDEVLVGYDR 60  
QY 58 VSAVPRDERFAVSREWESSAEYSAIPELSDMKKYGFLGFLPREDHARVKLVNPFSTR 117  
DB 61 AAVALADPRFS---KWRNS---TTPLEAALNNMLESPPRTTRKLVAREFTKR 114  
QY 118 AIDLRAEIQRTVDQLLDARSQOE--FDVVRDYAEGIPRAISALLKYPACDEKFRF 175  
DB 115 RVELLRPRVQELVDGLVDMMLAPDGRADLMESLAMPPLITYISLGVPEEDRAAFRW 174  
QY 176 GSATRALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRNPLENDVLTMLQAEAD 235  
DB 175 TDAF-----VFPPDPAQATAMAMSGYLSRL--IDSKRGQDEEDLSALVRTSD 224  
QY 226 GSRLSTKEIVALGAIAGDTTITYLIAFVNLNLSPEALELVAAEGLRNALDEV 295  
DB 225 GSRLTSEELGMAHILLVAGHETTVNLINMYALLSHPDQALALRADMTLLDGAVEEM 284  
QY 236 RFDNLIIRIGTVFAFADLEYCGASIKKGMVFLIPSLARDGTFSRPPVPRDGTAS 355  
DB 235 REGVESATVFPFEPVLDGTVLPAGDTVLVLADARTBERFPDPHRIIRDTAGH 344  
QY 356 LAYGCPHYCPGVSLARLEAEIAGTIFRRFPBMKLETP---VFGYPAFNRISLV 411  
DB 345 LAFGHGHPICAPLARLEARIAVRALRERCPDLADVSPGELWVYPMIRIGLKALPI 403

RESULT 10  
US-09-252-991A-17836  
Sequence 17836, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17836  
LENGTH: 437  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17836

Query Match 22.8%; Score 484.5; DB 4; Length 437;  
Best Local Similarity 33.5%; Pred. No. 2.3e-41;  
Matches 116; Conservative 74; Mismatches 139; Indels 17; Gaps 5;

QY 50 VWLTRHYDSAVFRDERFAVSREWESSAEYS---SAIPELSDMKKYGFLGFLPREDHAR 105  
DB 60 WVTVTRDARKVLYNH--GVRDARQAALVAKRTSPRAGIGEGSHMLNLDPPDHR 117  
QY 106 VAKLVNPFSTRPAILDLRAEIQRTVDQLLDARSQOEFPVVRDYAEGIPRAISALLKYP 165  
DB 118 IASLVGRATTPROVELOPHIERITBELLDAMAGRCADLMADFAIPLTAVIFELGLIP 177  
QY 166 AECDEKFRFSGATARALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRNPLENDV 225  
DB 178 EAEREHARQSWERQALLS-----PEERQALADQVDVLYRL--LEAKRKQPAD-DV 226  
QY 226 LTMLQAEADGSRSLSTKEIVALGAIAGDTTITYLIAFVNLNLSPEALELVAAE 285

DB 227 YSGLVQADDESGQISEAEIVSMAHLLMWSGFEITNMIGNLVTLVLPBEOALRAQDE 286  
QY 286 LMRNALDEVLRPDNLIIRIGTVFAFADLEYCGASIKKGMVFLIPSLARDGTFSRPPV 345  
DB 287 LIPNMBEIVRHDSVVRASMLRFTYEDVLDGVTTPAGEIIVSULTANHDAREDDPDR 346  
QY 346 FVVRDGTASLAVYGRPHVCPGVSLARLEAEIAGTIFRRFPBMKLE 391  
DB 347 LDTLNTDGHGIGYGVHYCVGASLARLEGRIALQRLARFPDQL 392

RESULT 11  
US-09-320-878-18  
Sequence 18, Application US/09320878A  
Patent No. 6117659  
GENERAL INFORMATION:  
APPLICANT: ASHLEY, Gary  
APPLICANT: BETLACH, Melanie C.  
APPLICANT: BETLACH, Mary C.  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 300622002120  
CURRENT APPLICATION NUMBER: US/09/320,878A  
CURRENT FILING DATE: 1999-05-27  
EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
EARLIER FILING DATE: 1998-08-28  
EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
EARLIER FILING DATE: 1998-05-06  
EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
EARLIER FILING DATE: 1997-04-30  
EARLIER APPLICATION NUMBER: 60/119,139  
EARLIER FILING DATE: 1999-02-08  
EARLIER APPLICATION NUMBER: 60/100,880  
EARLIER FILING DATE: 1998-09-22  
EARLIER APPLICATION NUMBER: 60/087,080  
EARLIER FILING DATE: 1998-05-28  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 18  
LENGTH: 416  
TYPE: PRT  
ORGANISM: Streptomyces venezuelae  
US-09-320-878-18

Query Match 22.8%; Score 484; DB 3; Length 416;  
Best Local Similarity 30.6%; Pred. No. 2.4e-41;  
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQSETPAPDFKFPAGYADDPFAIRLEATPI--FYWDEG-RSWLTRHYDSAV 61  
DB 5 QOGTTASPPVLDLGLGDFADPPYTAARLAEAGPAHRVTRPEDEVLVGYDRARAV 64  
QY 62 FDERFAVSREWESSAEYSAIPELSDMKKYGFLGFLPREDHARVKLVNPFSTRPIDL 121  
DB 65 LADPRFS---KWRNS---TTPLEAALNNMLESPPRTTRKLVAREFTKRVL 118  
QY 122 LRAEIQRTVDQLLDARSQOE--FDVVRDYAEGIPRAISALLKYPACDEKFRFPGAT 179  
DB 119 LRPVQELVDGLVDMMLAPDGRADLMESLAMPPLITYISLGVPEEDRAAFRWTD 178  
QY 180 AAALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRNPLENDVLTMLQAEADGSR 239  
DB 179 -----VFPPDPAQATAMAMSGYLSRL--IDSKRGQDEEDLSALVRTSDGSR 228  
QY 240 STKEIVALGAIAGDTTITYLIAFVNLNLSPEALELVAAEGLRNALDEVLRPDN 299  
DB 229 TSEELGMAHILLVAGHETTVNLINMYALLSHPDQALALRADMTLLDGAVEEMLYRG 288  
QY 300 IIRIGTVFAFADLEYCGASIKKGMVFLIPSLARDGTFSRPPVPRDGTASLAVG 359  
DB 289 PVESATVFPFEPVLDGTVLPAGDTVLVLADARTBERFPDPHRIIRDTAGHLARG 348

```

QY      360 RGHVCGVSLARLEAIEAVGTIFRRPEPKLKEP---VFGYHAFRNIESLNV 411
      349 HGHHFCIGAPLARLEAIEAVRALLERCPDLALDVSFGELVWYFNPIMRGKALPI 403

RESULT 12
US-09-141-908-13
Sequence 13, Application US/09141908
Patent No. 6503741
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
FILE REFERENCE: 30622002100
CURRENT APPLICATION NUMBER: US/09/141,908
CURRENT FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: PROV. 60/076,919
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: PROV. 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 416
TYPE: PR
ORGANISM: Streptomyces venezuelae
US-09-141-908-13

Query Match      22.8%; Score 484; DB 4; Length 416;
Best Local Similarity 30.6%; Pred. No. 2,4e-41;
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY      5 QANQSEKPAFPKPPAPGYADPPFAIEREATPI--FYWDEG-RSWLTRYHDVSAY 61
      5 QCGTTASPPVLIDGALGQDPADPYTYARLRABEGPAHVRTEGDEWLVVGYDARAV 64
      62 FRDERPAVREEMESSAEYSAPILSDMKYGLFGLPREDHARVRLVNPSTSAIDL 121
      65 LADPRS---KDMRNS---TTPLTEAEALNHNMLSDPPRHTRLRLVAREFTMRVEL 118
      122 LRAEIQRTVDQLDARSQGE--FDVVRDYAEGIPRAISALLKVAECDEKFRFGSAT 179
      119 LRPRVQEIYDGLVADNLAAPDGRADLMSLAWPLPTIVISELGVPEPDAARVWDAF 178
      180 ARAAGVGLVQVDEEKTIVASVTEGLALLHVDLDERRRNPLENDVLTMLQAEADGSR 239
      179 -----VFDDPAQQTAAEMSGYLSRL---IDSKRGDDGDELLSALVRTSDGSR 228
      240 STEELVALVCAITAAAGTDTTITLIAPAVNLNLSPEALEVKAEPGLMNADEVLRF 299
      229 TSEELGMMAHILLVAGHETTIVNLIANGMYALLSHPDQALALRDMTLLDGAVEEMLYEG 288
      300 ILRIGTRFARQDLEYCGASIKKGENVFLIISALRDGTVFSRPDVFDVRDGTASL 359
      289 PVSATYRFVEPEVDLDGTVIPAGDVLVVLADAHRTPERFPDHRDIRDTAGHLAF 348
      360 RGHVCGVSLARLEAIEAVGTIFRRPEPKLKEP---VFGYHAFRNIESLNV 411
      349 HGHHFCIGAPLARLEAIEAVRALLERCPDLALDVSFGELVWYFNPIMRGKALPI 403

RESULT 13
US-09-657-440-18
Sequence 18, Application US/09657440

```

```

Patent No. 6509455
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 30622002120
CURRENT APPLICATION NUMBER: US/09/657,440
CURRENT FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: 09/320,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: CIP OF 09/141,908
PRIOR FILING DATE: 1998-08-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 416
TYPE: PR
ORGANISM: Streptomyces venezuelae
US-09-657-440-18

Query Match      22.8%; Score 484; DB 4; Length 416;
Best Local Similarity 30.6%; Pred. No. 2,4e-41;
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY      5 QANQSEKPAFPKPPAPGYADPPFAIEREATPI--FYWDEG-RSWLTRYHDVSAY 61
      5 QCGTTASPPVLIDGALGQDPADPYTYARLRABEGPAHVRTEGDEWLVVGYDARAV 64
      62 FRDERPAVREEMESSAEYSAPILSDMKYGLFGLPREDHARVRLVNPSTSAIDL 121
      65 LADPRS---KDMRNS---TTPLTEAEALNHNMLSDPPRHTRLRLVAREFTMRVEL 118
      122 LRAEIQRTVDQLDARSQGE--FDVVRDYAEGIPRAISALLKVAECDEKFRFGSAT 179
      119 LRPRVQEIYDGLVADNLAAPDGRADLMSLAWPLPTIVISELGVPEPDAARVWDAF 178
      180 ARAAGVGLVQVDEEKTIVASVTEGLALLHVDLDERRRNPLENDVLTMLQAEADGSR 239
      179 -----VFDDPAQQTAAEMSGYLSRL---IDSKRGDDGDELLSALVRTSDGSR 228
      240 STEELVALVCAITAAAGTDTTITLIAPAVNLNLSPEALEVKAEPGLMNADEVLRF 299
      229 TSEELGMMAHILLVAGHETTIVNLIANGMYALLSHPDQALALRDMTLLDGAVEEMLYEG 288
      300 ILRIGTRFARQDLEYCGASIKKGENVFLIISALRDGTVFSRPDVFDVRDGTASL 359
      289 PVSATYRFVEPEVDLDGTVIPAGDVLVVLADAHRTPERFPDHRDIRDTAGHLAF 348
      360 RGHVCGVSLARLEAIEAVGTIFRRPEPKLKEP---VFGYHAFRNIESLNV 411
      349 HGHHFCIGAPLARLEAIEAVRALLERCPDLALDVSFGELVWYFNPIMRGKALPI 403

RESULT 14
5212296-9
Patent No. 5212296
APPLICANT: DEAN, CAROLINE HARDER, PATRICIA A.; LETO, KENNETH J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.; TEEPREMAN, JAMES M.
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING CYTOCHROMES
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/569,781
FILING DATE: 23-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 464,499
FILING DATE: 12-JAN-1990
APPLICATION NUMBER: 405,605
FILING DATE: 11-SEP-1989

```

SEQ ID NO:9:  
LENGTH: 403  
5212296-9

Query Match 21.6%; Score 459; DB 6; Length 403;  
Best Local Similarity 31.7%; Pred. No. 8.8e-39;  
Matches 127; Conservative 73; Mismatches 157; Indels 44; Gaps 12;

```

QY 21 AGVYEDFPFPAERLEATPI---FYWDEGRSVLTTRYHDSAVRDERDFVARSREWSS 77
DB 20 AFGCFDDPAFVDVTEAARTPEVTRATLMDGSCMLVTRHODVRAVGDPRF-----SA 71
QY 78 AEYSAPBELSDMKYKGFGLPP-----EDHAFVRKLVNPSFTSRAIDLRAEIQRTV 130
DB 72 DAHRTGFPLTAGR-EIIGTNPFTLRMDDEPHARLRRLTDLFVKKYEAARPEVQRLA 130
QY 131 DQILD-AASGGEFVVDYAGIEMRAISALLKTPACEDGKFRFGATARALGVGLVP 189
DB 131 DDLVDMRTGTRISADLVTEFALPLPSLVCLLGVPYE-DHAF--FOERSVLLTRSTP 187
QY 190 QVDEETKTLVASVTEGLALHDVLDERRNPLENDVLTMLQAEADGSRSTKELVALVG 249
DB 188 ---EVRRAQDELELYLARL---ARTKERPDALISRLVARGELDDTQIATMGRLLV- 240
QY 250 AIIAGDTTYIYLAFAVLNMLRSPEALELYAEBGLKRNALDEVLRFDNLIIGTVFA 309
DB 241 ---AGHETANMTALSTLVLRNPDQLARAEALVKGAVEELLRYLTIYHNGVPRLA 296
QY 310 RODLEYCGASIKGEMVFLIPSLRDGTVESRPVDFVRDPTGASLAYGSGPHVCPGV 369
DB 297 TEDVIGRTTIAAGGVCMISSANRDAEVFPGDDLVAARDARRHVAFGFVHQCLGQP 356
QY 370 LARLEAIIAVGTIFRRPPEMKL---KETP-----VFGYH 400
DB 357 LARVELQIAIETLRLRLPDRLAVPHEIIFRGMATYGVH 397

```

RESULT 15  
US-09-252-991A-32437  
; Sequence 32437, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32437  
; LENGTH: 468  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32437

Query Match 19.8%; Score 421.5; DB 4; Length 468;  
Best Local Similarity 29.5%; Pred. No. 8.3e-35;  
Matches 133; Conservative 70; Mismatches 177; Indels 71; Gaps 15;

```

QY 5 QANQSETKPA-----FDFKFPAGVADPPPAERLEATPIFYWDEGRSVLTTRYHD 57
DB 36 QPRHDAQRPALAPRSDGFDIHTYHPDVPALIRSRAPVCR-DQASIMWISRYAD 94
QY 58 VSAVFRDRERFAVS-----REEMESSAEYSASAIPELSDMKYKGLFGLPPEDHARVRK 108
DB 95 VSACLRDRRFSADPARLGAAGVRCGASWFGHQLOPLARFYDNFMLENDAPR-HTRLRR 153
QY 109 LVNPSFTSRAIDLLRAEIQRTVDQLDARSQGEFVDVADYAGIEMRAISALLKVPAC 168

```

```

DB 154 LPAPAFGPDVARRWEARIEVLVEELDSLERRRBDLDRFAEPLTIRVAELFGFPRED 213
QY 169 DEKPRRFGSATARALGVLPQVDEETKLVAS-----VTEGLALHDVDERRR- 219
DB 214 TGQLLPWG---RDLAAGL-----DLASHGDACQIRSAARFSDYLQROARGMSD 260
QY 220 -----PLENDVLTMLQAEADGSRSTKELVALVGAIIAAGTDTTYIYLAFAVLN 269
DB 261 GSRPRSGAASIFDGAAML---EAG---LGLLEDVLAAYAVFMAAFETITSMGNATLA 314
QY 270 LIRSPAELELYAEBGLRNALDEVLRFDNLIIGTVFAPADLEYCGASIKGEMVFL 329
DB 315 LITHPDQDLIRRCPELANAAYVEELLRPDGAVR-GGVACTLEEVYIGGQRIIPGGEKWL 373
QY 330 IPSALRDGTVESRPDVPDVR-DTGASLAYGSGPHVCGSVLARLEAIIAV-GTIFRRP- 386
DB 374 FLAARDEMPRAAPRLOLQORANAKOHVAFHAGPHYCLGAYLARLELOCALRGVRRFA 433
QY 387 -----PEMKKETPVGFGYHAPFRNIESLNV 412
DB 434 LASEPTDLRWRRSSV-----FRTLERLPiv 458

```

Search completed: October 2, 2003, 17:35:54  
Job time: 12.2473 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 28.237 Seconds

(without alignments)  
2355.298 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124  
Sequence: 1 MTOEQANQSEFKPAFDKRF.....HPAFRNISLVILKPSKAG 419

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_19Jun03.\*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2124	100.0	419	21	AA198579
2	528.5	24.8	396	21	AA198502
3	528.5	24.7	396	21	AA198503
4	518.5	24.4	396	21	AA198501
5	516	24.3	399	24	ABP57702
6	491	23.1	475	24	AAE34729
7	488.5	23.0	475	24	AAE34732
8	486.5	22.9	470	24	AAE34724
9	485.5	22.9	430	24	AAE34736

10	484	22.8	416	21	AA198654
11	484	22.8	416	21	AA198728
12	484	22.8	416	21	ABG71678
13	480.5	22.6	430	24	AAE34780
14	476.5	22.4	415	21	AA197196
15	475.5	22.4	430	24	AAE34725
16	475.5	22.4	430	24	AAE34727
17	475.5	22.4	430	24	AAE34731
18	466	21.9	432	24	AAE34726
19	463.5	21.8	430	24	AAE34737
20	452.5	21.3	430	24	AAE34730
21	452	21.3	482	23	ABB81316
22	451.5	21.3	429	24	AAE34728
23	450	21.2	396	23	ABB81315
24	449	21.1	448	24	ABU11985
25	448	21.1	430	24	AAE34723
26	448	21.1	430	24	AAE34735
27	447	21.0	430	24	AAE34733
28	443.5	20.9	430	24	AAE34734
29	443	20.9	399	24	ABU11376
30	435.5	20.5	426	24	AAE34738
31	435.5	20.5	430	24	AAE34722
32	432	20.3	403	12	AA111350
33	420	19.8	411	19	AAW54389
34	414.5	19.5	404	12	AA14724
35	413	19.4	396	15	AAE60777
36	407.5	19.2	587	18	AAW33274
37	406.5	19.1	407	24	ABP95990
38	405.5	19.1	418	21	AA193589
39	403.5	19.0	420	21	AA193790
40	401.5	18.9	402	24	ABU11981
41	386	18.2	398	22	AA1931560
42	385.5	18.1	405	24	ABP95987
43	382	18.0	406	24	ABP95992
44	371	17.5	409	22	AAE07915
45	370.5	17.4	398	17	AAW11585

## ALIGNMENTS

RESULT 1	
AA198579	standard; Protein, 419 AA.
ID	AA198579
AC	AA198579;
DT	10-APR-2000 (first entry)
DE	Sorangium cellulosum epothilone macrolactone oxidase EPOS F.
KM	EPOS F, epothilone macrolactone oxidase; epothilone biosynthesis; taxol substitute; anticancer.
OS	Sorangium cellulosum.
XX	
XX	MO9966028-A2.
PD	23-DEC-1999.
XX	
PF	16-JUN-1999; 99WO-EP04171.
XX	
PR	18-JUN-1998; 98US-0099504.
XX	24-SEP-1998; 98US-0101631.
PA	05-FEB-1999; 99US-0118906.
XX	
XX	(NOVS ) NOVARTIS AG.
XX	(NOVS ) NOVARTIS-ERFINDUNGEN VERN GRS MBH.
PI	Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX	WPI; 2000-097781/08.
DR	N-PSDB; AA255887.

Amino acid sequenc  
Macrolide ring mod  
S. venezuelae plic  
Streptomyces tuber  
S. venezuelae macr  
Streptomyces lydic  
Streptomyces chaib  
Streptomyces kasug  
Streptomyces speci  
Streptomyces lydic  
Streptomyces plate  
Streptomyces sp. c  
Streptomyces speci  
Streptomyces sp. c  
M. echinospora cal  
Streptomyces tuber  
Streptomyces plate  
Streptomyces tuber  
Streptomyces plate  
Protein encoded by  
Streptomyces tuber  
Streptomyces tuber  
Cytochrome enzyme  
Actinomadura hibi  
6-hydroxylase enco  
Mycinnamycin IV hyd  
S. fradiae tylosin  
Microtetraspora re  
Streptomyces globi  
S. fradiae tylosin  
M. echinospora cal  
Pimaricin biosynth  
Streptomyces sp. T  
Streptomyces livid  
S. clavuligerus pr  
Streptomyces prist

New isolated epothonione synthase genes, used for the recombinant production of epothonione for use in cancer therapy -

This sequence represents a *Sorangium cellulosum* epothilone macro-lactone oxidase, EPOS F, which is one of several epothilone biosynthetic enzymes encoded by a 60.75 kb contig. Epothilones A and B are 16-membered macrocyclic polyketides with an acylcysteine-derived starter unit; polyketides being synthesised from two-carbon building blocks, the beta-carbon of which always carries a keto group. Each round of two-carbon addition is carried out by a complex of enzymes known as the polyketide synthase in a manner similar to fatty acid biosynthesis. EPOS A (AAV585873) and EPOS P (AAV58574) are involved in formation of the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D and EPOS E (AAV58575-Y58578) are involved in polyketide backbone formation. EPO F (AAV58579) is an epothilone macro-lactone oxidase, and the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be involved in transport. Epothilones mimic the biological activity of taxol, and may be substituted for taxol in cancer chemotherapeutic compositions. Epothilones exhibit a much lower drop in potency against a multiply drug-resistant cell line compared with taxol, and are considerably less efficiently exported from such cells by the multidrug resistance protein (MDR, or P-glycoprotein). Despite the potential of epothilones as anticancer agents, they are problematical to produce on a large scale. Epothilones are too complex for industrial scale chemical synthesis, and *Sorangium cellulosum* is difficult to ferment, producing poor yields of epothilones. The nucleic acids of the invention may be used for the recombinant production of epothilones in a heterologous host that is more amenable to fermentation.

Q	Sequence	419 AA;
	Query Match	100.0%; Score 2124; DB 21;
	Best Local Similarity	100.0%; Pred. No. 6.8e-196; Length 419;
	Matches 419; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

Qy	61	MTQEOANQSTKCAFOFKEPAPGADPPALERLEATPIFYMEQGSWLTYYHDVSA	60
Db	1	MTQEOANQSTKCAFOFKEPAPGADPPALERLEATPIFYMEQGSWLTYYHDVSA	60
Qy	61	VFRDERPAVSREBESSAETSSAIPELSMKKYYGLGPEPDHAAVRKLVNPSTSAID	120
Db	61	VFRDERPAVSREBESSAETSSAIPELSMKKYYGLGPEPDHAAVRKLVNPSTSAID	120
Qy	121	LLBAELORTDOLLDRSGQEEFDVDRDAEGIPMRALSALLKYVAEDECDEFRRFGSATA	180
Db	121	LLBAELORTDOLLDRSGQEEFDVDRDAEGIPMRALSALLKYVAEDECDEFRRFGSATA	180
Qy	181	RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRLS	240
Db	181	RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRLS	240
Qy	241	TKELVALVGAIIAAGDTTITLYLAFANVLMILRSPEALBELVKAPEGMLNALDEVLRPDI	300
Db	241	TKELVALVGAIIAAGDTTITLYLAFANVLMILRSPEALBELVKAPEGMLNALDEVLRPDI	300
Qy	301	LRIGTVAFARODLEYGCASIKKGEWFFLLIPSLARDGYVFSRPDVFVDRDTGASLAYGR	360
Db	301	LRIGTVAFARODLEYGCASIKKGEWFFLLIPSLARDGYVFSRPDVFVDRDTGASLAYGR	360
Qy	361	GPHNCPCVSLARLEAELAVGTIFRRPEPKKLETPVGNHAFRNIESLANTILKPSKAG	419
Db	361	GPHNCPCVSLARLEAELAVGTIFRRPEPKKLETPVGNHAFRNIESLANTILKPSKAG	419

RESULT 2	
AAB15502	
ID	AAB15502 standard; Protein; 396 AA
XX	
AC	AAB15502;
XX	

DT 14-FEB-2001 (first entry)

**Bacillus subtilis hydroxylating protein #2.**

Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor; hydroxymethylglutaryl co-enzyme A reductase; hypocholesterolaemia

*Bacillus subtilis*.

WO200044886-A1.

03-AUG-2000.

28-JAN-2000; 2000WO-JP00472.

29-JAN-1999; 99JP-0021707.

( KYOW ) KYOWA HAKKO KOGYO KK.

Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;

WPI; 2000-548827/50.

1

New protein derived from *Bacillus* genus microorganism useful for producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase inhibitors -

Claim 10; Page 94-96; 11pp; Japanese.

This sequence represents a novel *Bacillus* derived protein having the activity of producing a hydroxylated bicyclic compound or the corresponding lactone from a bicyclic compound or the corresponding lactone. The protein is used for preparing hydroxylated compounds useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase inhibitors and hypcholesterolaemic agents.

**SQ Sequence 396 AA**

[illegible]

## RESULT 3

AAB15503  
ID AAB15503 standard; Protein; 396 AA.  
XX  
AC AAB15503;  
XX  
DT 14-FEB-2001 (first entry)  
XX  
DE Bacillus subtilis hydroxylating protein #3.  
XX  
KM Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;  
KM hydroxymethylglutaryl co-enzyme A reductase; hypocholsterolaemia.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200044886-A1.  
XX  
PD 03-AUG-2000.  
XX  
PF 28-JAN-2000; 2000WO-JP00472.  
XX  
PR 29-JAN-1999; 99JP-0021707.  
XX  
PA (KYOM) KYOMA HAKKO KOGYO KK.  
XX  
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
XX  
DR WPI; 2000-548827/50.  
XX  
DR N-PSDB; AAA95669.  
XX  
PT New protein derived from Bacillus genus microorganism useful for  
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase  
PS inhibitors -  
XX  
PS Claim 10; Page 105-107; 111pp; Japanese.  
XX  
CC This sequence represents a novel Bacillus derived protein having  
CC the activity of producing a hydroxylated bicyclic compound or the  
CC corresponding lactone from a bicyclic compound or the corresponding  
CC lactone. The protein is used for preparing hydroxylated compounds  
CC useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase  
CC inhibitors and hypocholsterolaemic agents.  
CC  
SQ Sequence 396 AA;  
Query Match 24.7%; Score 525.5; DB 21; Length 396;  
Best Local Similarity 33.2%; Pred. No. 3.2e-42; Indels 29; Gaps 6;  
Matches 126; Conservative 71; Mismatches 154;  
QY 25 AEDPPAERLRLEATPIFYWDEGRSVLTRYHDVSAVFRDERFAVSREEMESSAAYSSAI 84  
DB 23 AVHPFWYEMSKDAVPSFDEENQWVSFLYDDVKVGDKEI-----FSSCM 70  
QY 85 PELSMKKYGLFGLRPEDHARVKLVNPSFTSRALDLRAEIQRTYDQLDARSQGEED 144  
DB 71 PQOTSSIGNSIISMPPKTKIRSVNKAFTPRAMQWEPRIQETIDELIQFGRSERD 130  
QY 145 VVRDVAEGIPMRASIALTKVPAECDEKFRFSGSATARALGVGLVQVDEETTLVA---- 200  
DB 131 LVHDSYLPVIVISLGLVPSAHNEQFKMSD-----LVSTPKDXSEAEAKFLSEED 185  
QY 201 SVTEGIALLDVLDERRNPLENDVLTMLQAEADGSRISTKELVALVGAIITAGDTTIT 260  
DB 186 KCEELAAFPAGIIEKRNKPEODIISILVEAETGEKSGBELIPCTLLLVAGNETTT 245  
QY 261 YLIAAVNLRLSPALVLAEPGLMRNALDEVLRF---DNILRIGTVFRARODILEYCG 317  
DB 246 NLISNAMESILTEPGVYELRSHBELMPQAVEALRFAPAPVLR-----RIAKRDTETIGG 301  
QY 318 ASIKKGEMVFLIPSLRGTGFSRPDVFDRDTSALAYGSGPVCVCSLAREARI 377  
DB 302 HLIKEGDVTYLAVASANRDEAKFDRPHMDIRRHNPHTAFGHGHIIFCLGAPLARLEANI 361  
QY 378 AVGTIFRRPEMK-LKETPV 396

DB 362 ALTSLISAFPMHECVSITPI 381  
RESULT 4  
ID AAB15501 standard; Protein; 396 AA.  
XX  
AC AAB15501;  
XX  
DT 14-FEB-2001 (first entry)  
XX  
DE Bacillus subtilis hydroxylating protein.  
XX  
KM Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;  
KM hydroxymethylglutaryl co-enzyme A reductase; hypocholsterolaemia.  
XX  
OS Bacillus subtilis.  
XX  
PN WO200044886-A1.  
XX  
PD 03-AUG-2000.  
XX  
PF 28-JAN-2000; 2000WO-JP00472.  
XX  
PR 29-JAN-1999; 99JP-0021707.  
XX  
PA (KYOM) KYOMA HAKKO KOGYO KK.  
XX  
PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;  
XX  
DR WPI; 2000-548827/50.  
XX  
DR N-PSDB; AAA95669.  
XX  
PT New protein derived from Bacillus genus microorganism useful for  
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase  
PS inhibitors -  
XX  
PS Claim 8; Page 68-71; 111pp; Japanese.  
XX  
CC This sequence represents a novel Bacillus derived protein having  
CC the activity of producing a hydroxylated bicyclic compound or the  
CC corresponding lactone from a bicyclic compound or the corresponding  
CC lactone. The protein is used for preparing hydroxylated compounds  
CC useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase  
CC inhibitors and hypocholsterolaemic agents.  
CC  
SQ Sequence 396 AA;  
Query Match 24.4%; Score 518.5; DB 21; Length 396;  
Best Local Similarity 32.6%; Pred. No. 1.5e-41;  
Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;  
QY 25 AEDPPAERLRLEATPIFYWDEGRSVLTRYHDVSAVFRDERFAVSREEMESSAAYSSAI 84  
DB 23 AVHPFWYEMSKDAVPSFDEENQWVSFLYDDVKVGDKEI-----FSSCM 70  
QY 85 PELSMKKYGLFGLRPEDHARVKLVNPSFTSRALDLRAEIQRTYDQLDARSQGEED 144  
DB 71 PQOTSSIGNSIISMPPKTKIRSVNKAFTPRAMQWEPRIQETIDELIQFGRSERD 130  
QY 145 VVRDVAEGIPMRASIALTKVPAECDEKFRFSGSATARALGVGLVQVDEETTLVA---- 200  
DB 131 LVHDSYLPVIVISLGLVPSAHNEQFKMSD-----LVSTPKDXSEAEAKFLSEED 185  
QY 201 SVTEGIALLDVLDERRNPLENDVLTMLQAEADGSRISTKELVALVGAIITAGDTTIT 260  
DB 186 KCEELAAFPAGIIEKRNKPEODIISILVEAETGEKSGBELIPCTLLLVAGNETTT 245  
QY 261 YLIAAVNLRLSPALVLAEPGLMRNALDEVLRF---DNILRIGTVFRARODILEYCG 317  
DB 246 NLISNAMESILTEPGVYELRSHBELMPQAVEALRFAPAPVLR-----RIAKRDTETIGG 301





QY	48	RMVLTIRHYDVSAVFPEDEEPFAVSRREWMESSAEYSSAIPETLS-----DMKKYGGFGJP---	99
Db	94	PIMVLTIRFPDVAEVWADQRF-----VNNPTIVFGIGADQDPRAKLIETLFGILEDL	143
QY	100	-----PEDHARVRKVLNVPSEFTSRALDILRAEIQRTVDOLIDARSGOE--FD	144
Db	144	AEYLTDTLITSDPPDHTRLRLRLVRSAPFTARRIQDLPRVRERTIDELLARIPDHAEDGVVD	203
QY	145	VVRDVAEGIPMAAISLLKVPACDEKFRFRFGSATRALGVGVPOVDEETKLVASYTE	204
Db	204	LVENHVAVPLPIVVICELVGIDEDRALMRFFGADLA-----SLNP-----KRIGATMPE	252
QY	205	GLALLHDVDERRNRNLENDVLVLMLOA--EADSRSLSTKELVAVGAIIAAGTDTTYLI	263
Db	253	MLAHIHVEYDERRAD--LRDDLLSGLLRADDDGRLSDVEMVTLVTLVLAGHETTHLI	311
QY	264	AFAVLNLRSPALAEIVKAPGGLMRNALDEVLFEDNILRIGTVRFARQDLEYCGASIKG	323
Db	312	SNGTALLTLTRPDQRRLIDEDPALLPRAVHELMKWCSPIQATQRLRYAMEDTEVAGVVRQG	371
QY	324	ENVFLLIPBALDGVTFSRPDVDFVDRDTC-----SLAYGRGHVCPGVSLALEAEIAY	379
Db	372	EALMFSLVLANHDPRHYTGPERLDLITRQPAAGRAEDHVFGHGMYCCIGASLARQEAEEVAY	431
QY	380	GTIFRRFPPEMKLETTP	395
Db	432	GKLTLLRYPDLALATLP	447
RESULT 7			
AAE34732			
ID	AAE34732	standard; Protein; 475 AA.	
AC	AAE34732;		
XX	DT	14-MAY-2003 (first entry)	
XX	XX	Streptomyces rimosus email protein.	
DE	XX		
XX	XX	P450 monooxygenase; avermectin; ferredoxin; ferredoxin reductase; enzyme;	
KW	XX	emamectin; insecticide; email protein.	
XX	OS	Streptomyces rimosus.	
PN	WO200292801-A2.		
PD	21-NOV-2002.		
XX	XX		
XX	PF	15-MAY-2002; 2002WO-EP05363.	
XX	PR	16-MAY-2001; 2001US-291149P.	
XX	PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.	
XX	PI	Molnar I, Ligson JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;	
XX	PI	Buckel TG;	
DR	WPI: 2003-140280/13.		
XX	DR	N-PSDB; AADS3027.	
PT	Novel polypeptide that exhibits enzymatic activity of P450		
PT	monooxygenase and capable of regioselectively oxidizing alcohol of		
CC	avermectin useful for making emamectin from avermectin -		
CC	Claim 17; Page 121-122; 157pp; English.		
CC	The present invention relates to novel proteins that exhibit an enzymatic		
CC	activity of P450 monooxygenase and capable of regioselectively oxidizing		
CC	the 4-carbonyl group at position 4 of avermectin to 4-keto-avermectin.		
CC	The invention also relates to ferredoxins and ferredoxin reductases that		
CC	are active with the P450 monooxygenases. Sequences of the invention are		
CC	useful for producing 4-keto-avermectin from avermectin, which is useful		

CC	for producing emamectin. Emamectin is useful as an insecticide. The
CC	present sequence is Streptomyces rimosus emm1 protein.
XX	
sq	Sequence 475 AA;
Query Match	23.0%; Score 488.5; DB 24; Length 475;
Best Local Similarity	29.7%; Pred. No. 1.7e-38;
Matches 131;	Conservative 83; Mismatches 158; Indels 69; Gaps 12
QY	2 TOEOANOSETK-PADEFKPAFGVA-----EDFPFAIEREATPIF--- 42
DB	29 TPSAATSSDITYPATTDRTLLPSYGLHGPBEPNLMPELIDNPYGYGTLRQAPLVAR 88
QY	43 YWDEKRSVLLRHHVSAVFRDERAVAREEWESSAEVSSAIPELSDMK-----YGLG 97
DB	89 FIDOSPVLVTRFDVREVMEDQRF-----VNNPTLVPGIGADKDPRARLIELFG 138
QY	98 LP-----PEDHARVKKLVNPSFTSRAIDLLRAEIORVQDLDAKSGQEE 142
DB	139 IPEDLITPLADITILTSDPEPDHTRLRRLVSRATARIQDLRRVREDITALLERLPDHE 198
QY	143 --FDVDVDAEGIPMRAISALLKVPACDEKFRFRGATARALGVLVPQVDEETKTLV 199
DB	199 DQVVDLVHEFPAVPLPIITVICELVGIDEDRDTLMRRFGADLA-----SLNP-----KRIG 247
QY	200 ASVTBGLALLDVLDERRRNPENLVLTMLQA-EADGSRSLTKEIVLVGAIIAAGTDT 258
DB	248 ATPEMIMIHIEVIDERRA-ALRDLLEGLIRAODDGGRLSDVENVTLLTVLAGHET 306
QY	259 TIYIAFVNLILNSPPALLETVKAEPGLMRNALDEVLRPDNLTIRIGTVAFARQDLEYCGA 318
DB	307 TAILLSNCTLALLTHPDORRLIDEDPALLPRVHELMRWCGPIQATQTLRYAMEDETVAGV 366
QY	319 SIKKGEWFLIIPSAIRDGVFSRDPDVRDRTGA---SLAYGRGPHVCEGVSLARLE 374
DB	367 QVRGEALMFLVAVANHDPRHHTGTGERLDLTRQAPGRABDHVGFHGMYCIGASLARQE 426
QY	375 AEIAVGTIFRRPPEMKLKEPT 395
DB	427 AEVAVGKLLTRYPDLELATP 447
RESULT 8	
AAE34724	
ID	AAE34724 standard; Protein; 470 AA.
AC	AAE34724;
XX	
DT	14-MAY-2003 (first entry)
XX	
DE	Streptomyces rimosus emm3 protein.
XX	
XX	P450 monooxygenase; avermectin, ferriodoxin; ferriodoxin reductase; enzyme;
KM	emamectin; insecticide; emm3 protein.
OS	Streptomyces rimosus.
XX	
PN	MO200292801-A2.
PD	21-NOV-2002.
XX	
PF	15-MAY-2002; 2002WO-EP05363.
XX	
PR	16-MAY-2001; 2001US-291149P.
XX	
PA	(SYGN ) SYNGENTA PARTICIPATIONS AG.
XX	
PI	Mojnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JB;
XX	Buckel TG;
XX	WPI; 2003-140280/13.
DR	N-PSDB; AAD53019.
XX	

PT Novel polypeptide that exhibits enzymatic activity of P450  
 PT monooxygenase and capable of regioselectively oxidizing alcohol of  
 PT avermectin useful for making emamectin from avermectin -  
 XX  
 XX  
 PS Claim 17; Page 107-108; 157pp; English.

CC The present invention relates to novel proteins that exhibit an enzymatic  
 CC activity of P450 monooxygenase and capable of regioselectively oxidizing  
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.  
 CC The invention also relates to ferredoxins and ferredoxin reductases that  
 CC are active with the P450 monooxygenases. Sequences of the invention are  
 CC useful for producing 4-keto-avermectin from avermectin, which is useful  
 CC for producing emamectin. Emamectin is useful as an insecticide. The  
 CC present sequence is Streptomyces rimosus emas protein.

XX  
 XX  
 SQ Sequence 470 AA;

Query Match 22.9%; Score 486.5; DB 24; Length 470;  
 Best Local Similarity 30.6%; Pred. No. 2.6e-38;  
 Matches 124; Conservative 81; Mismatches 147; Indels 53; Gaps 10;

QY 22 PGAAEDPPAIELEARETPIF---YWDGRSVLTRYHDVSAVFRDERAVSRREMESSA 78  
 DB 60 PELLEPPYGYGTLRQAPLVRRARFIDDSPIMLVTRFDVVRVREVRDQRF-----V 109

QY 79 EYSSAIPBELSDMKK---YGLFGLP-----PDHARVRLVNPSTFSA 118  
 DB 110 NNNTLVPGICADDDPRARLIEFGIPEDLAPYLTDNLITSDPDHTRRLVSRATFARR 169

QY 119 IDLLRAEIQRTVDQLDARSQGE---FDVVRDYAGCIPMRASALLKVPACDEKFRFP 175  
 DB 170 IQLRPRVERITDELRLRPHADGVLDVEHFAVPLPTVLCELVGIDEEDRALMRRF 229

QY 176 GSATAALGVLPQVDEETKTIVASTEGALLHVDLDRRRNPLENDVLTMLQA-EA 234  
 DB 230 GAOLA-----SLNP-----KRIGATVPEMISHIHETIDERRA-ALRDDLGLIRQDD 277

QY 235 DGSRLSTKELVALVGAIIAAGTDTTYLAFVNLNLSPEALELVKAEFGMLRNALDEV 294  
 DB 278 DGSRLSDVENVTLVTLVLAGHETTAHLINGTALLTHPDQRLLIDEDPALPRAVHEL 337

QY 295 LRFENLIRIGTVAFARODLEYCGASIKGEMVFLIPSAIRDGTFSRPDPVRRDTGA 354  
 DB 338 MRWCGPIQATQLRYALDETEVAGVQVQGBALMFSLVAAHNDPRHYTEPRRLDTIQPAG 397

QY 355 ----SLAYGRPHVCGVSIARLEAEIANGVTIRRRPEMKLXETP 395  
 DB 398 RABDHVGFQGHMHCIGASLARQBAEVAVYKLLTRYPDALALTLP 442

RESULT 9  
 AAE34736  
 ID AAE34736 standard; Protein; 430 AA.

XX  
 XX  
 AC AAE34736;  
 XX  
 DT 14-MAY-2003 (first entry)

DE Streptomyces lydicus emas5 protein.  
 XX  
 KW P450 monooxygenase; avermectin; ferredoxin; ferredoxin reductase; enzyme;  
 KW emamectin; insecticide; emas5 protein.

XX  
 OS Streptomyces lydicus.  
 XX  
 PN WO200292801-A2.  
 XX  
 PD 21-NOV-2002.

XX  
 PF 15-MAY-2002; 2002WO-EP05363.  
 XX  
 PR 16-MAY-2001; 2001US-291149P.  
 XX

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;  
 XX PI Buckel TG;

XX  
 DR WPI; 2003-140280/13.  
 DR N-PSDB; AAD530311.

XX  
 PT Novel polypeptide that exhibits enzymatic activity of P450  
 PT monooxygenase and capable of regioselectively oxidizing alcohol of  
 PT avermectin useful for making emamectin from avermectin -  
 XX  
 XX  
 PS Claim 17; Page 128-129; 157pp; English.

CC The present invention relates to novel proteins that exhibit an enzymatic  
 CC activity of P450 monooxygenase and capable of regioselectively oxidizing  
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.  
 CC The invention also relates to ferredoxins and ferredoxin reductases that  
 CC are active with the P450 monooxygenases. Sequences of the invention are  
 CC useful for producing 4-keto-avermectin from avermectin, which is useful  
 CC for producing emamectin. Emamectin is useful as an insecticide. The  
 CC present sequence is Streptomyces lydicus emas5 protein.

XX  
 XX  
 SQ Sequence 430 AA;

Query Match 22.9%; Score 485.5; DB 24; Length 430;  
 Best Local Similarity 31.4%; Pred. No. 2.9e-38;  
 Matches 129; Conservative 81; Mismatches 150; Indels 51; Gaps 13;

QY 22 PGAAEDPPAIELEARETPIF---YWDGRSVLTRYHDVSAVFRDERAVSRREMESSA 78  
 DB 25 PALIGDPFAGYALREOGFVVRGRFVDDSPVWFVTRFEEVREVLDRQRF---RNPPVSSA 81

QY 79 EYSSAIPBELSDMKK-YGLFGLP-----PDHARVRLVNPSTFSAIDL 122  
 DB 82 --PDADPEPTPLSRIMDMGFPENLRVYLGSILNNDADPDHTRRLVSRATFARKITDL 139

QY 123 RAEIQRTVDQLDARSQGE---FDVVRDYAGCIPMRASALLKVPACDEKFRFRGSAT 179  
 DB 140 RRVQAQIABELLARLEPHNEDGVLDIQHFAVPLPTVLCELVGIPEDRPQWRMGADL 199

QY 180 APALGVLPQVDEETKTIVASTEGALLHVDLDRRRNPLENDVLTMLQA-EADGR 238  
 DB 200 ----VSLQP---DRMSRSPAMIDH---IHELIAARR-ALTDLLSLITHTDDGGR 247

QY 239 LSTKEVALVGAIIAAGTDTTYLAFVNLNLSPEALELVKAEFGMLRNALDEVLRD 298  
 DB 248 LSDVENVTLVTLVLAGHETTAHLINGTALLTHPDQRLLIDEDPALPRAVHELMRWC 307

QY 299 NILRIGTVAFARODLEYCGASIKGEMVFLIPSAIRDGTFSRPDPVRRDTGA---- 354  
 DB 308 GPVHMTQLRYAEDVGLAVIRKGDVQGLIVSANRDRHYTEPRRLDTIRPAGHAEN 367

QY 355 SLAYGRPHVCGVSIARLEAEIANGVTIRRRPEMK-----LKETPFG 398  
 DB 368 HVGFGHGAHYCIGATLAKQEGVALGALLRHFPSELAVAPDALERTVPVG 418

RESULT 10  
 AAB18654  
 ID AAB18654 standard; Protein; 416 AA.

XX  
 XX  
 AC AAB18654;  
 XX  
 DT 22-JUN-2001 (first entry)

DE Amino acid sequence of picromycin/methymycin cytochrome P450 hydroxylase.

XX  
 KW Nartonolide synthase; polyketide synthase gene; nartonolide polyketide;  
 KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;  
 KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;  
 KW picromycin biosynthesis;  
 KW picromycin/methymycin cytochrome P450 hydroxylase.

```

XX OS Streptomyces venezuelae.
XX PN US6117659-A.
XX PD 12-SEP-2000.
XX PF 27-MAY-1999; 99US-0320878.
XX PR 28-MAY-1998; 98US-0087080.
XX PR 22-SEP-1998; 98US-0100880.
XX PR 08-FEB-1999; 99US-0119139.
XX PR 20-MAY-1999; 99US-0134990.
XX PR 30-APR-1997; 97US-0846247.
XX PR 06-MAY-1998; 98US-0073538.
XX PR 28-AUG-1998; 98US-0141908.
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX DR WPI; 2000-610844/58.
XX PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
XX PT for converting ketolides to antibiotics and as antibiotics and
XX PT intermediates in the synthesis of compounds with pharmaceutical value
XX PS Disclosure; Columns 47-48; 117pp; English.
XX CC The present sequence represents a picromycin/methymycin cytochrome P450
XX CC hydroxylase (pick). The nucleotide sequence encoding it is used in the
XX CC course of the invention. The specification describes a recombinant DNA
XX CC compound expressing recombinant polyketide synthase genes in host cells
XX CC for the production of narbonolide, narbonolide derivatives and
XX CC polyketides that are useful as antibiotics and as intermediates in the
XX CC synthesis of compounds with pharmaceutical value. The DNA compounds may
XX CC also encode a C12-hydroxylase (pick), desosamine biosynthesis and
XX CC desosaminyl transferase enzymes (useful for conversion of ketolides to
XX CC antibiotics), and the beta-glucosidase enzyme (involved in picromycin
XX CC biosynthesis). These compounds are also useful for increasing the
XX CC antibiotic activity of a compound relative to the unhydroxylated
XX CC compound. The recombinant host cells are useful as genetic systems that
XX CC allow rapid engineering of the narbonolide polyketide synthase. These
XX CC would be valuable for creating novel ketolide analogs for pharmaceutical
XX CC applications.
XX SQ Sequence 416 AA;
XX
XX Query Match 22.8%; Score 484; DB 21; Length 416;
XX Best Local Similarity 30.6%; Pred. No. 3.8e-38;
XX Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;
XX
XX 5 QANGETKPAFDFKPPAGYADPPAIBRLREAPPI--FYWDEG-RSWVLTTRYHDVSAV 61
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 5 OCGTASFPVLVLGALGQDFADPYTYARLAEAGAHVRPPEGDEVWLTVGYDRAV 64
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 62 FDEDFAVSREWEESAIESSAIPESIDMKKYGFLPEPDARVAKLVNPFSTRADL 121
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 65 IADPFRS--KDWRS--TTPLTEEALNNNMESDPPRTTRKTLVAREFTMRKRVL 118
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 122 LRAEIORTVQDLLDRSGQE--FVVVRDYAEGIPRAISALLKVPACDEKFRFGSAT 179
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 119 LRPRQDEIVDGLVDMALAPDGRADLMESLAMPPTTVISELLGVPEPRRAFRVWTDNF 178
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 180 ARAAGVGLVQVDEETKTLVASVTEGLALLHVDLERRRNPENDVLTMLQAEDGSR 239
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 179 -----VPPDDPAQQTMAEMSGYLSRL--IDSKRGQDGEDLLSALVRSDEDSRL 228
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 240 STKEIVLVGATIAAGTDTTTLIAVAVNLRSPPALVVAEPMANALDEVIRFNP 299
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 229 TSEILLGMHILLVAGHETTVNLIAWGVALLSHPDQLAALRADMTLLDGAVEEMLRYG 288
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX

```

```

QY 300 IIRIGTVPARODEYCCASIKKGMVFLIPSAIRDGTVERPDPVDRDGTASLAVG 359
DB 289 PVESATYFPPVVDLGTVPAGDTVLVLADARTRPERFPDPFRDRIARRDTACHLAFG 348
QY 360 RGPVHCPSVSLARLAEIAVGTIFRRPFEMLKEFP---VPGYHAFPNIESLVN 411
DB 349 HGIFHCIGAPLARLEARIARLALRCPDLALDVSFGELWVYPMIRGLKALPI 403

RESULT 11
AA67218
ID AA67218 standard; protein; 416 AA.
XX
XX AA67218;
AC 23-MAR-2000 (first entry)
XX
XX DE Macrolide ring modifier pick hydroxylase amino acid sequence.
XX KW Narbonolide polyketide synthase; PKS; hydroxylase; pick; ketolide;
XX KW antibiotic production; narbomycin; picromycin.
XX OS Streptomyces venezuelae.
XX PN WO9961599-A2.
XX PD 02-DEC-1999.
XX PF 27-MAY-1999; 99WO-US11814.
XX PR 28-MAY-1998; 98US-0087080.
XX PR 28-AUG-1998; 98US-0141908.
XX PR 22-SEP-1998; 98US-0100880.
XX PR 08-FEB-1999; 99US-0119139.
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX DR WPI; 2000-072618/05.
XX DR N-PSDB; AA256001, AA256003.
XX PT New recombinant DNA encoding a domain of narbonolide polyketide
XX PT synthase, for production of ketolide antibiotics -
XX PS Example 2; Page 40; 98pp; English.
XX
XX CC This is the Streptomyces venezuelae hydroxylase pick protein sequence.
XX CC The pick protein is involved in macrolide ring modification. The
XX CC invention relates to recombinant DNA containing a coding sequence for a
XX CC narbonolide polyketide synthase (PKS). Polyketides are compounds
XX CC synthesised from 2-carbon units through a series of condensations and
XX CC subsequent modifications. Modular PKSs are responsible for the
XX CC production of many antibiotics including picromycin. The narbonolide PKS
XX CC consists of a loading module, six extender modules, and two thioester
XX CC domains. Four proteins make up the narbonolide PKS (PICAI, PICAIL,
XX CC PICAIIT and PICAIIV). PICAI includes the loading module and extender
XX CC modules 1 and 2, PICAIL includes extender modules 3 and 4, PICAIIT
XX CC includes extender module 5 and PICAIIV includes extender module 6 and a
XX CC type II thioesterase domain. The second type II thioesterase domain is
XX CC found on the PICB protein. The nucleotide sequences encoding all of these
XX CC proteins can be isolated in recombinant form from the recombinant cosmid
XX CC PKO5023-27 (see AA256001). Narbonolide is desosaminylated in S.
XX CC venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
XX CC required for this conversion, and the desosamine biosynthetic genes are
XX CC also found in cosmid PKO5023-27. The recombinant DNA of the invention is
XX CC used to express, in transformed cells, narbonolide (or its derivatives)
XX CC or other ketolides (particularly hybrids), which may then be converted
XX CC (e.g. by other enzymes recombinantly expressed in the same hosts) to
XX CC polyketide antibiotics or their intermediates. The antibiotics are useful
XX CC in human or veterinary medicine.
XX SQ Sequence 416 AA;
XX

```



XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;  
 PI Buckel TG;  
 XX WPI; 2003-140280/13.  
 DR N-PSDB; AAD53052.  
 XX Novel polypeptide that exhibits enzymatic activity of P450  
 PT monooxygenase and capable of regioselectively oxidizing alcohol of  
 PT avermectin useful for making emamectin from avermectin -  
 XX  
 PS Claim 17; Page 147-148; 157pp; English.  
 XX  
 CC The present invention relates to novel proteins that exhibit an enzymatic  
 CC activity of P450 monooxygenase and capable of regioselectively oxidizing  
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.  
 CC The invention also relates to ferredoxins and ferredoxin reductases that  
 CC are active with the P450 monooxygenases. Sequences of the invention are  
 CC useful for producing 4-keto-avermectin from avermectin, which is useful  
 CC for producing emamectin. Emamectin is useful as an insecticide. The  
 CC present sequence is Streptomyces tubercidicus emat7 protein.

XX Sequence 430 AA;

Query Match 22.6%; Score 480.5; DB 24; Length 430;

Best Local Similarity 29.1%; Pred. No. 8,8e-38; Mismatches 161; Indels 67; Gaps 12;

14 AEDFKPFA-----PGVADEPPAIELEAREATPIF--YWDGGRSWLTRYH 56  
 3 AISSSPFAHVGHKPGEPVMDPALITDPFGGALREGGYPVLRGFMDSPLVLTRE 62  
 57 DVSAVFRDRFAVSRREWESSAAYSALPELSDMKYGLFGLPP-----101  
 63 EVROVLRDGRFLNN---PAASPGHSIDESPRLALDMGMEHFRPYLMSILNNDAP 118  
 102 DHARVKLVNPSPTSAIDLLRAEIQRTVDOLLANSQGE---FVYVADYAGIPMAI 158  
 119 DHTLRRLVSRATARKITDLRPRVQLADELRLRPEHAGDVVLLIHFAIPFTVI 178  
 159 SALLKVPACDEKFRFRGSATRALGVLPQVDEETKTLVASVTEGLALLHDVDERRR 218  
 179 CELVGPEADRPQWRKMG-----DLVSLQPELSTSPFAMIEH---IHLEIRE-RR 226  
 219 NPLENDVLTMLIA--EADGSRLSTKELVALGAIAGTDTTYLITAFVNLILRSPAL 277  
 227 GALTDLLELIRTHDDGSRSLSDVEMVTMVLVLAGHETTHALIGNSTAALLTRPD 286  
 278 ELVKAEPGLMRNALDVAREFNILRIGTVRFARODLEYGASIKKEMFLLIPSLRGG 337  
 287 RLVKDPELLPRVNHLLRMCSPVQOTQRYASEDEVEIGTPRKGDVAQLIVSANFDP 346  
 338 TVFSRPDVVRDRTGA---SLAYGRPHVCPGVSLARLEAIVAGTIFRRPEM-----389  
 347 RHYTAPERDLTRHPAGHAENHVGFGHNYCGALILAQOEGVAGKLTHTYBELSLAV 406  
 390 ---KKETPVFGYHPRFNIESINVL 413  
 407 APDELERTPV---PGSWRLDSLPLVRL 429

RESULT 14

AA77196 standard; Protein; 415 AA.

AA77196;

05-JUN-2000 (first entry)

S. venezuelae macrolide biosynthetic enzyme pikC, SEQ ID NO:39.

XX

KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
 KW chronic obstructive pulmonary disease; respiratory inflammation;  
 KW hypercholesterolemia; crop protection agent.

XX Streptomyces venezuelae ATCC15439.

XX Key Location/Qualifiers

XX Misc-difference 103 /note= "Encoded by CCG"

XX WO200000620-A2.

XX 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MINU ) UNIV MINNESOTA.

XX Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI; 2000-160679/14.

XX N-PSDB; AA287301.

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

XX synthesis of methymycin and pikromycin -

XX Claim 19; Page 429-430; 438pp; English.

XX The invention relates to an isolated and purified nucleic acid segment  
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
 CC biologically active variant, where the nucleic acid sequence is not  
 CC derived from the eryc gene cluster of Saccharopolyspora erythraea or  
 CC streptomycetes antibiotics. The invention also relates to a macrolide  
 CC biosynthetic gene cluster, or fragments thereof. The macrolide  
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,  
 CC pikromycin, neomethymycin, narbomycin or a combination of these  
 CC compounds. Recombinant or augmented cells comprising the desosamine  
 CC and/or macrolide biosynthetic gene clusters are useful for the  
 CC production of biologically active macrolides. The macrolide biosynthetic  
 CC proteins are useful for synthesis of methymycin, pikromycin,  
 CC neomethymycin and narbomycin. The alternative termination of polyketide  
 CC synthesis may be useful to prepare novel antibiotics and  
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the  
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or  
 CC biomedical applications, to engineer PHA monomer synthases or to prepare  
 CC biologically active agents, such as chemotherapeutics.  
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary  
 CC disease as well as other diseases involving respiratory inflammation,  
 CC cholesterol-lowering agents or macrolide-based antibiotics which are  
 CC active against a variety of organisms, e.g., bacteria, including  
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well  
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,  
 CC fungicides or insecticides) via expression of polyketides in plants.  
 CC Sequences AA77190-77197 represent macrolide biosynthetic enzymes from  
 CC Streptomyces venezuelae ATCC 15439, which are encoded by sequences  
 CC AA287295-287302.

XX Sequence 415 AA;

Query Match 22.4%; Score 476.5; DB 21; Length 415;

Best Local Similarity 30.5%; Pred. No. 2e-37; Mismatches 185; Indels 25; Gaps 9;

1 MTOBOANOSETKAPDFKPFAGYADPEPPAIELEAREATPI--FYWDG-BSWVLTTRYND 57

1 MRRTOGGTTASPPVLDLALGODPRAADYPTTARLARAGPARRVTPREDDVMTLVGICVR 60

58 VSAVFRDRFAVSRREWESSAAYSALPELSDMKYGLFGLPPEDHARVKLVNDSFTSR 117

```

Db      61 ARAVLADPPRES---KOWRNS---TTPLTAEALNHNMLESDPPRHTR-RKLVAEFTMR 113
Qy      118 AITLLAEIORTVDLLDASGOEE--FDVVRDYAEGIPKRAISALLKVAECDEKRRF 175
      114 RVLLLPFRVGEIYDGLVDALAAPDGRADLMESLAMPITVISELIGVEPPRAARVW 173
Qy      176 GSATARALGVLPOVDEETKTLVASVTEGLALHLDVDERRRNPLENDVLTMLQAEAD 235
      174 TDF-----VPPDDPAQOQTMAEMSGYLSRL---IDSKRGQDGDGLSALVRSDED 223
Db      236 GSRLLSTKEVLVALGAIIAAGTDTTIVLIAFAVLNLRSPALBLVKAEPGLMRNALDEV 295
      224 GSRLLTSEELGMHIIILVAGHEFTVLIANGMYALLSHPPDLAALRADMTLLDGAVEML 283
Qy      296 RPNILRIGTVRPARODLEYCGASIKKGEVFPILLPSALRDGTVEFSRDPVEDVRDTCAS 355
      284 RYGPVESATYRRPVEPVDLDGTVPAGDVLVLADAHRTPERFPDPHFDLRDPAH 343
Db      356 IAYGRGPHVCPGVSILAEIAVGTIFRRFPPEMKLKEP---VFGYHFAFRNIESLNV 411
      344 LAFGHGHCIGAPLARLEARIARVALLERCPDLADVSGELVWYFNPIMRIGLKALPI 402

RESULT 15
ID AAE34725 standard; Protein; 430 AA.
AC AAE34725;
XX
XX 14-MAY-2003 (first entry)
DE Streptomyces lydicus ema4 protein.
XX
XX P450 monooxygenase; avermectin; ferredoxin; ferredoxin reductase; enzyme;
KM emamectin; insecticide; ema4 protein.
XX
XX Streptomyces lydicus.
OS
XX
XX WQ200292801-A2.
PN
XX 21-NOV-2002.
PD
XX 15-MAY-2002; 2002MO-EP05363.
PF
XX 16-MAY-2001; 2001US-291149P.
PR
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA
XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
XX WPI; 2003-140280/13.
DR N-PSDB; AAD53020.
XX
XX Novel polypeptide that exhibits enzymatic activity of P450
PT monooxygenase and capable of regioselectively oxidizing alcohol of
PT avermectin useful for making emamectin from avermectin -
XX
XX Claim 17; Page 109-110; 157pp; English.
XX
XX The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidizing
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferredoxins and ferredoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces lydicus ema4 protein.
XX
SQ Sequence 430 AA;

```

Query Match 22.4%; Score 475.5; DB 24; Length 430;  
 Best Local Similarity 30.9%; Pred. No. 2.7e-37;

```

Matches 127; Conservative 81; Mismatches 152; Indels 51; Gaps 13;
Qy      22 PGYAEDEPFAIRLREATPIF---YVDEGRSWLTYHDVSAVFRERPAVSREEMESSA 78
      25 PALIGDPFVGYGALRQGVVVKGRFNDSPWFVTRFEVREYVLRDPRF---RNNVSA 81
Db      79 EVSSAIPELSDMKK-YGLFGLPPE-----DHARVRKLVNPSFTSRAIDLL 122
      82 --PGAAPEDTPLSRLLMDMGFPPEHLRVLLGSIINNDADPHTRRLRVSRAPFAFKITDL 139
Qy      123 RAEIQTVDQLDARSQEE---FDVVRDYAEGIPKRAISALLKVAECDEKRRRGSAT 179
      140 RPRVTOIADBLARLEBEHAEAGVVDLIQFAFPLPTVICELVGIPEEDRPQWRWGADU 199
Qy      180 ARAVGVLPOVDEETKTLVASVTEGLALHLDVDERRRNPLENDVLTMLQAEADGSR 238
      200 ----VSLDP--DRMSRSFPAMIDH---IHELIAARR-ALTDDLSELIRTHDDGSR 247
Qy      239 LSTKELVALGAIIAAGTDTTIVLIAFAVLNLRSPALBLVKAEPGLMRNALDEVLRPD 298
      248 LSDVENVTWLVTLVLAHETTAHLIGNTAALLTHPDQLRLKDPALLPRAVHELMRMC 307
Db      299 NILRIGTVRPARODLEYCGASIKKGEVFPILLPSALRDGTVEFSRDPVEDVRDTCG 354
      308 GPVHMTQLRYAAEDVELAGVRIKRTGDAVOLIVSANRDRPHYDLPDRLDTRHPAGHAEN 367
Qy      355 SLAYGRGPHVCPGVSILAEIAVGTIFRRFPPEMK-----LKETPVFG 398
      368 HVGFGHAHYCLGATLAKOGEVALGALRHFEELSIAVAPEALENTPTVPG 418
Db

```

Search completed: October 2, 2003, 17:33:53  
 Job time : 30.237 secs